DOCUMENT RESUME

ED 378 624	CS 508 800
AUTHOR	Fowler, Carol A., Ed.
TITLE	Status Report on Speech Research, July-December 1993. SR-115/116.
INSTITUTION	Haskins Labs., New Haven, Conn.
SPONS AGENCY	National Inst. of Child Health and Human Development (NIH), Bethesda, MD.
PUB DATE	93
CONTRACT	DBS-9112198; HD-01994; HD-21888
NOTE	162p.; For the April-June 1993 report, see ED 371 423.
PUB TYPE	Collected Works - General (020) -- Reports - Research/Technical (143)
EDRS PRICE	MF01/PC07 Plus Postage.
DESCRIPTORS	Adults; Communication Research; Elementary Secondary Education; Higher Education; Infants; *Language Acquisition; Language Research; *Listening Skills; Literacy; Phonology; *Speech Communication
IDENTIFIERS	*Speech Research

ABSTRACT

This publication (one of a series) contains 12 articles which report the status and progress of studies on the nature of speech, instruments for its investigation, and practical applications. Articles in the publication are: "Dynamics and Coordinate Systems on Skilled Sensorimotor Activity" (Elliot L. Saltzman); "Speech Motor Coordination and Control: Evidence from Lip, Jaw, and Laryngeal Movements" (Vincent L. Gracco and Anders Lofqvist); "An Unsupervised Method for Learning to Track Tongue Position from an Acoustic Signal" (John Hogden and others); "Prosodic Patterns in the Coordination of Vowel and Consonant Gestures" (Caroline L. Smith); "Divergent Developmental Patterns for Infants' Perception of Two Non-Native Consonant Contrasts" (Catherine T. Best and others); "Beyond Orthography and Phonology: Differences between Inflections and Derivations" (Laurie Beth Feldman); "Visual and Phonological Determinants of Misreadings in a Transparent Orthography" (G. Cossu and others); "Phonological Computation and Missing Vowels: Mapping Lexical Involvement in Reading" (Ram Frost); "The Tritone Paradox and the Pitch Range of the Speaking Voice: A Dubious Connection" (Bruno H. Repp); "A Review of Treiman, R. (1993). 'Beginning to Spell'" (Donald Shankweiler); "A Review of McNeill, D. (1992). 'Hand and Mind: What Gestures Reveal about Thought'" (Michael Studdert-Kennedy); and "A Review of Lieberman, P. (1991). 'Uniquely Human'" (Michael Studdert-Kennedy). (RS)

****
*   Reproductions supplied by EDRS are the best that can be made   *
*   from the original document.   *
****

ED 378 624

CS

Haskins Laboratories Status Report on

Speech Research

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

- Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

SR-115/116
JULY-DECEMBER 1993

CS508800

BEST COPY AVAILABLE

Haskins
Laboratories
Status Report on
Speech Research

SR-115/116
JULY-DECEMBER 1993

NEW HAVEN, CONNECTICUT

Distribution Statement

Editor
Carol A. Fowler

Production
Yvonne Manning-Jones
Fawn Zefang Wang

This publication reports the status and progress of studies on the nature of speech, instrumentation for its investigation, and practical applications.

Distribution of this document is unlimited. The information in this document is available to the general public. Haskins Laboratories distributes it primarily for library use. Copies are available from the National Technical Information Service or the ERIC Document Reproduction Service. See the Appendix for order numbers of previous Status Reports.

Correspondence concerning this report should be addressed to the Editor at the address below:

Haskins Laboratories
270 Crown Street
New Haven, Connecticut
06511-6695

Phone: (203) 865-6163 *FAX:* (203) 865-8963 *Bitnet:* HASKINS@YALEHASK
Internet: HASKINS%YALEHASK@VENUS.YCC.YALE.EDU

This Report was reproduced on recycled paper

Acknowledgment

The research reported here was made possible in part by support from the following sources:

National Institute of Child Health and Human Development

Grant HD-01994
Grant HD-21888

National Institute of Health

Biomedical Research Support Grant RR-05596

National Science Foundation

Grant DBS-9112198

National Institute on Deafness and Other Communication Disorders

Grant DC 00121
Grant DC 00183
Grant DC 00403
Grant DC 00016
Grant DC 00594
Grant DC 00865
Grant DC 01147
Grant DC 00044
Grant DC 00825
Grant DC 01247

Investigators

Arthur S. Abramson*
Peter J. Alfonso*
Eric Bateson*
Fredericka Bell-Berti*
Shlomo Bentin*
Catherine T. Best*
Susan Brady*
Catherine P. Browman
Claudia Carello*
Franklin S. Cooper*
Stephen Crain*
Lois G. Dreyer*
Alice Faber
Laurie B. Feldman*
Janet Fodor*
Anne Fowler*
Carol A. Fowler*
Ram Frost*
Louis Goldstein*
Carol Gracco
Vincent Gracco
Katherine S. Harris*
Leonard Katz*
Rena Arens Krakow*
Andrea G. Levitt*
Alvin M. Liberman*
Diane Lillo-Martin*
Leigh Lisker*
Anders Löfqvist
Georgije Lukatela*
Ignatius G. Mattingly*
Nancy S. McGarr*
Richard S. McGowan
Weijia Ni
Patrick W. Nye
Kiyoshi Oshima†
Kenneth Pugh*
Lawrence J. Raphael*
Bruno H. Repp
Hyla Rubin*
Philip E. Rubin
Elliot Saltzman
Donald Shankweiler*
Jeffrey Shaw
Michael Studdert-Kennedy*
Michael T. Turvey*
Douglas Whalen

Technical Staff

Michael D'Angelo
Vincent Gulisano
Donald Hailey
Yvonne Manning-Jones
William P. Scully
Fawn Zefang Wang
Edward R. Wiley

Administrative Staff

Philip Chagnon
Alice Dadourian
Betty J. DeLise
Lisa Fresa
Joan C. Martinez

Students*

Melanie Campbell
Sandra Chiang
Terri Erwin
Douglas Honorof
Pai-Ling Hsiao
Laura Koenig
Simon Levy
Subhobrata Mitra
Mira Peter
Joaquin Romero
Dorothy Ross
Arlyne Russo
Michelle Sancier
Sonya Sheffert
Brenda Stone
Mark Tiede
Qi Wang

*Part-time
†Visiting from University of Tokyo, Japan

Contents

Dynamics and Coordinate Systems in Skilled Sensorimotor Activity
Elliot L. Saltzman 1

Speech Motor Coordination and Control: Evidence From Lip, Jaw, and Laryngeal Movements
Vincent L. Gracco and Anders Löfqvist 17

An Unsupervised Method for Learning to Track Tongue Position from an Acoustic Signal
John Hogden, Philip Rubin, and Elliot Saltzman 33

Prosodic Patterns in the Coordination of Vowel and Consonant Gestures
Caroline L. Smith 45

Divergent Developmental Patterns for Infants' Perception of Two Non-Native Consonant Contrasts
Catherine T. Best, Gerald W. McRoberts, Rosemarie LaFleur, and Jean Silver-Isenstadt 57

Beyond Orthography and Phonology: Differences between Inflections and Derivations
Laurie Beth Feldman 69

Visual and Phonological Determinants of Misreadings in a Transparent Orthography
G. Cossu, D. P. Shankweiler, I. Y. Liberman, and M. Gugliotta 99

Phonological Computation and Missing Vowels: Mapping Lexical Involvement in Reading
Ram Frost 113

The Tritone Paradox and the Pitch Range of the Speaking Voice: A Dubious Connection
Bruno H. Repp 127

A Review of Treiman, R. (1993). *Beginning to Spell*
Donald Shankweiler 145

A Review of McNeill, D. (1992). *Hand and Mind: What Gestures Reveal About Thought*
Michael Studdert-Kennedy 149

A Review of Lieberman, P. (1991). *Uniquely Human*
Michael Studdert-Kennedy 155

Appendix 157

7

Haskins
Laboratories
Status Report on
Speech Research

Haskins Laboratories Status Report on Speech Research
1993, SR-115/116, 1-15

Dynamics and Coordinate Systems in Skilled Sensorimotor Activity*

Elliot L. Saltzman†

1. INTRODUCTION

Skilled sensorimotor activities entail the creation of complex *kinematic* patterns by actors using their limbs and speech articulators. Examples of kinematic patterns include trajectories over time of a reaching hand's position, velocity, or acceleration variables, the spatial shape of the path taken by a hand-held pen during handwriting, or the relative timing of the speech articulators to produce the phonemes /p/, /e/, and /n/ in the word "pen". The term *dynamics* is used to refer to the vector field of forces that underlies and gives rise to an action's observable kinematic patterns. In this chapter, a dynamical account of skilled activity is reviewed in which skilled behavior is characterized as much as possible as that of a relatively autonomous, self-organizing dynamical system. In such systems, task-appropriate kinematics are viewed as emerging from the system's underlying dynamical organization (e.g., Beek, 1989; Saltzman & Munhall, 1989; Schöner & Kelso, 1988; Turvey, 1990). Thus, the emphasis of the present account is on a dynamical description, rather than a kinematic one, of sensorimotor skills. For example, an extreme and admittedly exaggerated "straw man" counter-hypothesis is that of a central executive or homunculus that produces a given movement pattern with reference to an internal kinematic template of the form, tracing out the form provided by the template, and using the articulators as a physiological and biomechanical pantograph to produce a larger version of the pattern in the external world.

An adequate account of skilled sensorimotor behaviors must also address the multiplicity of coordinate systems or state spaces, and the mappings or transformations that exist among them, that appear to be useful in describing such behaviors. For example, a reaching movement can be described simultaneously in terms of patterns of muscle activations, joint angle changes, spatial motions of the hand, etc., and in terms of the ways these patterns relate to one another. This chapter focuses on the roles of both dynamics and coordinate systems in skilled sensorimotor activities. Evidence is reviewed in this chapter supporting the claim that the dynamics of sensorimotor control and coordination are defined in highly abstract coordinate systems called *task spaces* that are distinct from, yet related to, the relatively concrete physiological and biomechanical details of the peripheral musculoskeletal apparatus. It is further hypothesized that such spaces are the media through which actions are coupled perceptually to task-relevant surfaces, objects, and events in the actor's environment.

The chapter is divided into roughly two parts. The first is focused on concepts of dynamics as they have been applied to understanding the performance of single or dual sensorimotor tasks, where each task is defined in a one-to-one manner with a single articulatory degree of freedom. For example, a single task could be defined as the oscillation of a hand about the wrist joint or of the forearm about the elbow joint; a dual task could be defined as the simultaneous oscillations of both the right and left hand, or of the elbow and hand of a given arm. The second part of the chapter is focused on how the notions of dynamics and

---

This work was supported by grant support from the following sources: NIH Grant #DC-00121 (Dynamics of Speech Articulation) and NSF Grant #BNS-88-20099 (Phonetic Structure Using Articulatory Dynamics) to Haskins Laboratories. I am grateful to Claudia Carello, Philip Rubin, and Michael Turvey for helpful comments on earlier versions of this chapter.

coordinate systems can be combined or synthesized to account for the performance of single or multiple tasks, where each task is defined over an entire effector system with many articulatory degrees of freedom. For example, in the production of speech the task of bringing the lips together to create a bilabial closure for /p/ is accomplished using the upper lip, lower lip, and jaw as articulatory degrees of freedom.

2. Dynamics

Why place so much emphasis on the dynamics of sensorimotor coordination and control? A dynamical account of the generation of movement patterns is to be preferred over other accounts, in particular the notion of internal kinematic templates, because dynamics gives a unified and parsimonious account of (at least) four signature properties of such patterns:

(1) *Spatiotemporal form*. A movement's spatiotemporal form can be described both qualitatively and quantitatively. For example, qualitatively different hand motions are displayed in situations where the hand moves discretely to a target position and then stops, and where the hand moves in a continuous, rhythmic fashion between two targets. Quantitative differences are reflected in the durations and extents of various discrete motions, and in the frequencies and amplitudes of the rhythmic motions;

(2) *Stability*. A movement's form can remain stable in the face of unforeseen perturbations to the state of the system encountered during movement performances;

(3) *Scaling*. Lawful warping of a movement's form can occur with parametric changes along performance dimensions such as motion rate and extent; and

(4) *Invariance and variability*. A dynamical framework allows one to characterize in a rigorous manner a common intuition concerning skilled actions in general. This intuition is that there is a subtle underlying invariance of control despite an obvious surface variability in performance.

In order to illustrate these points, the behavior of several simple classes of dynamical systems will be reviewed (e.g., Abraham & Shaw, 1982; Baker & Gollub, 1990; Thompson & Stewart, 1986; see also Norton, in press). Mathematical models based on these systems have been used to provide accounts of and to simulate the performance of simple tasks in the laboratory. In such models, the qualitative aspects of a system's dynamics are mapped onto the functional characteristics of the performed tasks. For example, discrete positioning tasks can be modeled as being governed globally by *point attractor* or *fixed point* dynamics. Such dynamical systems move from initial states in a given neighborhood, or *attractor basin*, of an attracting point to the point itself in a time-asymptotic manner. Similarly, sustained oscillatory tasks can be modeled using *periodic attractor* or *limit cycle* dynamics. Such dynamics move systems from initial states in the attractor basin of an attracting cycle to the cycle itself in a time-asymptotic manner (see Examples 8 and 9 in Norton [in press] for representative equations of motion and sets of state trajectories for fixed point and limit cycle systems, respectively). The performance of simultaneous rhythms by different effectors can be modeled as the behavior of a system of *coupled* limit cycle oscillators, in which the motion equation of each oscillator includes coupling term(s) that represent the influence of the other oscillator's ongoing state. For example, the coupling term in oscillator-*i*'s equation of motion might be a simple linear function, $a_{ij}x_j$, of the position of oscillator-*j*, where x_j is the ongoing position of oscillator-*j* and a_{ij} is a constant coefficient that maps this position into a coupling influence on oscillator-*i*. In what follows, discussion is focused initially on single degree-of-freedom oscillatory tasks, and then moves to comparable, dual degree-of-freedom tasks.

Single Degree-of-freedom Rhythms

In a typical single degree-of-freedom rhythmic task, a subject is asked to produce a sustained oscillatory movement about a single articulatory degree of freedom, e.g., of the hand or a hand-held pendulum about the wrist joint. Usually, the rhythm is performed at either a self-selected "comfortable" frequency or at a frequency specified externally by metronome; in both cases, the amplitudes of the performed oscillations are self-selected according to comfort criteria. Such movements can be characterized as limit cycle oscillations, in that they exhibit characteristic frequencies and amplitudes (e.g., Kugler & Turvey, 1987) that are stable to externally imposed perturbations (e.g., Kay, Saltzman, & Kelso, 1991; Scholz & Kelso, 1989). For example, after such rhythms are subjected to brief mechanical perturbations, they return spontaneously to their original pre-perturbation frequencies and amplitudes. Additionally, limit cycle models capture the spontaneous covariation or scaling behavior that is observed among the task's kinematic observables. For example, at a given movement frequency there is a highly linear relationship between a cycle's motion amplitude and it's peak velocity, such that cycles with larger amplitudes generally display greater peak velocities. Such a relationship is inherent in the dynamics of near-sinusoidal limit cycle oscillations. Further, across a series of different metronome-specified frequencies, the mean cycle amplitude decreases systematically as cycle frequency increases (e.g., Kay, Kelso, Saltzman, & Schöner, 1987). Such scaling is a natural consequence of the structure of the limit cycle's *escapement*, a nonlinear damping mechanism that is responsible for offsetting frictional losses and for governing energy flows through the system in a manner that creates and sustains the limit cycle's rhythm.

Dual Degree-of-freedom Rhythms

These tasks consist simply of two single degree-of-freedom tasks performed simultaneously, e.g., rhythmic motions of the right and left index fingers, usually at a common self-selected or metronome-specified frequency and with self-selected amplitudes. Additionally, subjects are requested typically to perform the task with a given relative phasing between the component rhythms (e.g., Kelso, 1984; Rosenblum & Turvey, 1988; Sternad, Turvey, & Schmidt, 1992; Turvey & Carello, in press). For example, for bimanual pendulum oscillations performed at a common frequency in the right and left parasagittal planes (see Figure 7, Turvey & Carello, in press), an *in-phase* relationship is defined by same-direction movements of the components, i.e., front-back movements of the right pendulum synchronous with front-back movements of the left pendulum; similarly, an *antiphase* relationship is defined by simultaneous, opposite-direction movements of the components. Models of such tasks begin by specifying each component unit as a separate limit cycle oscillator, with a 1:1 frequency ratio defined between the pair of oscillators. If this were all there was to the matter, one could create arbitrary phase relations between the component limit cycles, simply by starting the components with an initial phase difference equal to the desired phase difference. This is an inadequate description of dual rhythmic performances, however, since the behavioral data demonstrate that it is only possible to easily perform 1:1 rhythms that are close to inphase or antiphase; intermediate phase differences are not impossible, but they require a good deal of practice and usually remain more variable than the inphase and antiphase pair.

What makes the inphase and antiphase patterns so easy to perform, and the others so difficult? What is the source of this natural cooperativity? It turns out that these are the same questions that arise when one considers the phenomenon of *entrainment* between limit cycle oscillators. This phenomenon was observed by the 17th century Dutch physicist Christian Huygens, who noticed that the pendulum swings of clocks placed on the same wall tended to become synchronized with one another after a period of time. This phenomenon can be modeled dynamically by assuming that each clock is its own limit cycle oscillator, and that the clocks are coupled to one another due to weak vibrations transmitted through the wall. Such coupling causes the motions of the clocks to mutually perturb one another's ongoing rhythms, and to settle into a cooperative state of entrainment. These observations suggest that the appropriate theory for understanding the performance of multiple task rhythms is that of coupled limit cycle oscillators. In this theory, when two limit cycles are coupled bidirectionally to one another, the system's behavior is usually attracted to one of two *modal* states. In each modal state, the components oscillate at a common mode-specific frequency, and with a characteristic amplitude ratio and relative phase. Most important for the present discussion, if the component oscillators are roughly identical and the coupling strengths are roughly the same in both directions, then the two modes are characterized by relative phases close to inphase and antiphase, respectively. It is possible, however, that the frequencies and amplitudes observed in the modal states can be different from those observed when the components oscillate independently of one another.

Thus, we are led to view the inphase and antiphase coordinative patterns in 1:1 dual oscillatory tasks as the attractive modal states of a system of coupled limit cycle components. Note that the coupling that creates this modal cooperativity is involuntary and obligatory, in the sense that these modal states are hard to avoid even if the task is to perform with a relative phasing in between those of the naturally easy modes. Such intermediate states are possible to perform, but require much practice and remain more variable than the modal states. What is the structure of the intercomponent coupling? What is the source or medium through which this coupling is defined?

Coupling structure. Coupling structure refers to the mathematical structure of the coupling functions that map the ongoing states of a given oscillator into perturbing influences on another. It turns out that many types of coupling will create stable modes with relative phases close to inphase and antiphase. For example, even the simple linear positional coupling mentioned earlier, $a_{ij}x_j$, will work, where x_j is the ongoing position of oscillator-j and a_{ij} is a constant coefficient that maps this position into a perturbation of oscillator-i's motion.

In addition to entrainment, however, human rhythmic tasks display *phase transition* behaviors that place additional constraints on the choice of coupling functions. In an experimental paradigm pioneered by Kelso (e.g., Kelso, 1984; Scholz & Kelso, 1989), subjects begin an experimental trial by oscillating two limb segments at the same frequency in an antiphase pattern, and then increase the frequency of oscillation over the course of the trial. Under such conditions, the antiphase coordination abruptly shifts to an inphase coordination when the oscillation frequency passes a certain critical value. A comparable shift is not seen, however, when subjects begin with an inphase pattern; under these conditions, the inphase coordination is maintained as frequency increases. The abrupt phase transition from antiphase to inphase patterns when frequency is increased can be characterized mathematically as a *bifurcation* phenomenon in the underlying dynamical system. In dynamical models of such phenomena the coupling functions are required typically to be nonlinear (e.g., Haken, Kelso, & Bunz, 1985; Schöner, Haken, & Kelso, 1986). To summarize briefly, entrainment can be created by limit cycles coupled bidirectionally in many ways, but entrainment with bifurcations require typically nonlinear coupling structures.

Coupling medium. What is the source of interoscillator coupling during the performance of simultaneous rhythmic tasks? What are the coordinates along which such coupling is defined? One possibility is that the coupling medium is mechanical in nature, as in the case of Huygens' pendulum clocks, since it is known that biomechanical *reactive coupling* exists among the segments of effector systems during motor skill performances (e.g., Bernstein 1967/1984; Hollerbach, 1982; Saltzman, 1979; Schneider, Zernicke, Schmidt, & Hart, 1989). Such coupling is defined in segmental or joint-space coordinate systems. A second possibility is that the coupling is neuroanatomical, as in the case of the crosstalk or overflow between neural regions controlling homologous muscle groups that has been hypothesized to underly mirroring errors in bimanual sequencing tasks such as typing or key-pressing (e.g., MacKay & Soderberg, 1971), or associated mirror movements in certain clinical populations (e.g., Woods & Teuber, 1978). Such coupling is defined in muscle-based coordinate systems.

An experiment by Schmidt, Carello, and Turvey (1990) indicated that matters might not be so straightforward. In this experiment, subjects performed rhythmic motions at their knee joints, but the major innovation of the paradigm was to have the set of two rhythms defined across subjects rather than within subjects. Thus, one subject would perform rhythmic oscillations at one knee joint while watching a nearby partner do the same (see Figure 9, Turvey & Carello, in press). There were two types of task. In one type, the partners were asked to oscillate their respective legs at a mutually comfortable common frequency either inphase or antiphase with one another, and to increase or decrease the oscillation frequency by self-selected amounts in response to a signal supplied by the experimenter; in the second type of task, a metronome was used to specify both the frequencies and time schedule of frequency scaling. Surprisingly, all the details of entrainment and bifurcation phenomena were observed in this between-person experiment as had been observed previously in the within-person experiments. Clearly, joint-space (biomechanical) and muscle-space (neural) coordinates were not the media of interoscillator coupling in this experiment. Rather, the coupling must have been due to visual *information* that was specific to the observed oscillatory states of the pendulums themselves. The same point has received further support in subsequent studies in which similar behaviors are displayed by subjects who oscillate an index finger either on or off the beat provided auditorily by metronome (Kelso, Delcolle, & Schöner, 1990), or who oscillate a forearm inphase or antiphase with the visible motion of a cursor on a CRT screen (van Riel, Beek, & van Wieringen, 1991). All these studies underscore the conclusion that the coupling medium is an abstract one, and that coupling functions are defined by perceptual information that is specific to the tasks being performed.

Coordinative dynamics. Just as the coupling medium is not defined in simple anatomical or biomechanical terms, several lines of evidence support the hypothesis that the limit cycle dynamics themselves are also not specified in this manner. That is, the degrees-of-freedom or state variables along which the oscillatory dynamics are specified, and that experience the effects of interoscillator coupling, are not defined in simple anatomical or biomechanical coordinates. Even tasks that, at first glance, might appear to be specified at the level of so-called "articulatory" joint rotational degrees-of-freedom have been found to be more appropriately characterized in terms of the orientations of body segments in body-spatial or environment-spatial coordinate systems. For example, Baldissera, Cavallari, and Civaschi (1982) studied the performance of simultaneous 1:1 oscillations about the ipsilateral wrist and ankle joints in the parasagittal plane. Foot motion consisted of alternating downward (plantar) and upward (dorsal) motion. Hand motion consisted of alternating flexion and extension. The relationship between anatomical and spatial hand motions was manipulated across conditions by instructing subjects to keep the forearm either palm down (pronated) or palm up (supinated). Thus, anatomical flexion/extension at the wrist caused the hand to rotate spatially downward/upward during the pronation condition, but spatially upward/downward during supination. It was found that the easiest and most stably performed combinations of hand and foot movements were those in which the hand and foot motions were in the same spatial direction, regardless of the relative phasing between upper and lower limb muscle groups. Thus, the easiest and most natural patterns were those in which hand and foot motions were spatially inphase. It was more difficult to perform the spatially antiphase combinations, and occasional spontaneous transitions were observed from the spatially antiphase patterns to the spatially inphase patterns. Related findings on combinations of upper and lower limb rhythmic tasks were more recently reported by Baldissera, Cavallari, Marini, and Tassone (1991) and by Kelso and Jeka (1992).[1]

Thus, the dynamical systems for coordination and control of sensorimotor tasks, and the medium through which these systems are coupled, cannot be described in simple biomechanical or neuroanatomical terms. Rather, they are defined in abstract, spatial, and informational terms. This point becomes even clearer when one examines the performance of tasks that are more realistic and complex than the relatively artificial and simple tasks that have been reviewed above.

Speech Production

Consider the production of speech and what is entailed during the speech gesture of raising the tongue tip toward the roof of the mouth to create and release a constriction for the phoneme /z/, using the tongue tip, tongue body and jaw in a synergistic manner to attain the phonetic goal. Such systems show a remarkable flexibility in reaching such task goals, and can compensate adaptively for disturbances or perturbations encountered by one part of the system by spontaneously readjusting the activity of other parts of the system in order to still achieve these goals. An elegant demonstration of this ability was provided in an experiment by Kelso, Tuller, Vatikiotis-Bateson, and Fowler (1984; see also Abbs & Gracco, 1983; Folkins & Abbs, 1975; Shaiman, 1989). In this experiment, subjects were asked to produce the syllables /bæb/ or /bæz/ in the carrier phrase "It's a ___ again", while recording (among other observables) the kinematics of upper lip, lower lip, and jaw motion, as well as the electromyographic activity of the tongue-raising genioglossus muscle. During the experiment, the subjects' jaws were unexpectably and unpredictably perturbed downward as they were moving into the final /b/ closure for /bæb/ or the final /z/ constriction for /bæz/. It was found that when the target was /b/, for which lip but not tongue activity is crucial, there was remote compensation in the upper lip relative to unperturbed control trials, but normal tongue activity (Figure 1A); when the target was /z/, for which tongue but not lip activity is crucial, remote compensation occurred in the tongue but not the upper lip (Figure 1B). Furthermore, the compensation was relatively immediate in that it took approximately 20-30 ms from the onset of the downward jaw perturbation to the onset of the remote compensatory activity. The speed of this response implies that there is some sort of automatic "reflexive" organization established among the articulators with a relatively fast loop time. However, the gestural specificity implies that the mappings from perturbing inputs to compensatory outputs is not hard-wired. Rather, these data imply the existence of a task- or gesture-specific, selective pattern of coupling among the component articulators that is specific to the utterance or phoneme produced.

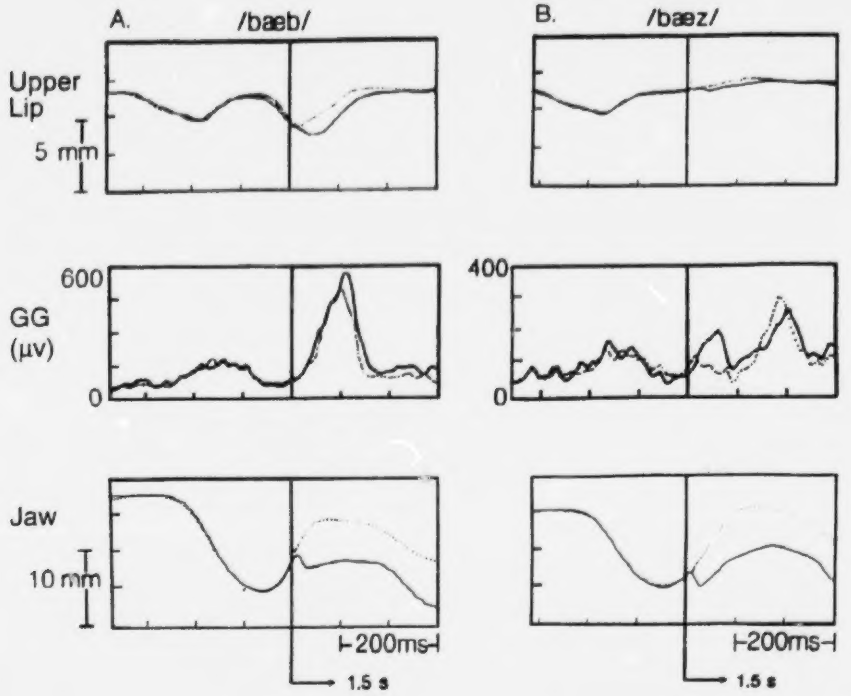


Figure 1. Experimental trajectory data for the unperturbed (dotted lines) and perturbed (solid lines) utterances /bæb/ (column A) and /bæz/ (column B). Top row: Upper lip position; Middle row: Genioglossus muscle activity; Bottom row: Jaw position; Panels in each column are aligned with reference to the perturbation onset (Solid vertical lines); Perturbation duration was 1.5 s. (adapted from Figures 3 & 5 in "Functionally specific articulatory cooperation following jaw perturbations during speech: Evidence for coordinative structures" by J. A. S. Kelso, B. Tuller, E. Vatikiotis-Bateson, & C. A. Fowler, 1984).

What kind of dynamical system can display this sort of flexibility? Clearly, it cannot be a system in which task goals are defined independently at the level of the individual articulators. For example, if one were to model a bilabial closing gesture by giving each articulatory component (upper lip, lower lip, and jaw) point attractor dynamics and its own target position, then the system would attain a canonical closure in unperturbed simulations. However, the system would fail in simulations in which perturbing forces were added to one of the articulators during the closing gesture. For example, if a simulated braking force were added to the jaw that prevented it from reaching its target, then the overall closure goal would not be met even though the remaining articulators were able to attain their own individual targets.

Appropriately flexible system behavior can be obtained, however, if the task-specific dynamics are defined in coordinates more abstract than those defined by the articulatory degrees-of-freedom. Recall that, in earlier discussions of coupled limit cycle dynamics, the term "modal state" was used to characterize the cooperative states that emerged from the dynamics of the coupled system components. Modal patterns defined the systems' preferred or natural set of behaviors. The problem at hand, therefore, is to understand how to create modal behaviors that are tailored to the demands of tasks encountered in the real world. This can be accomplished if one can design task-specific coupling functions among a set of articulatory components that serve to create an appropriate set of task-specific system modes. The remainder of this chapter will be devoted to describing one approach to the design of task-specific dynamical systems, called *task-dynamics*, that has been used with some success to model the dynamics of speech production. This modeling work has been performed in cooperation with several colleagues at Haskins Laboratories as part of an ongoing project focused on the development of a gesturally-based, computational model of linguistic structures (e.g., Browman & Goldstein, 1986, 1991, in press; Fowler, & Saltzman, 1993; Kelso, Saltzman, & Tuller, 1986a, 1986b; Kelso, Vatikiotis-Bateson, Saltzman, & Kay, 1985; Saltzman, 1986, 1991; Saltzman & Kelso, 1987; Saltzman & Munhall, 1989). For recent reviews, related work, and critiques, see also de Jong (1991), Edwards, Beckman, and Fletcher (1991), Hawkins (1992), Jordan and Rosenbaum (1989), Mattingly, (1990), Perkell (1991), and Vatikiotis-Bateson (1988).

3. TASK DYNAMICS

The discussion of task dynamics for speech production is divided into two parts. The first focuses on the dynamics of interarticulatory coordination within single speech gestures, e.g., the coordination of lips and jaw to produce a bilabial closure. The second part focuses on the dynamics of intergestural coordination, with special attention being paid to periods of *coproduction* when the blended influences of several temporally overlapping gestures are evident in the ongoing articulatory and acoustic patterns of speech (e.g., Bell-Berti & Harris, 1981; Fowler, 1980; Fowler & Saltzman, 1993; Harris, 1984; Keating, 1985; Kent & Minifie, 1977; Öhman, 1966, 1967; Perkell, 1969; Sussman, MacNeilage, & Hanson, 1973). For example, in a vowel-consonant-vowel (VCV) sequence, much evidence supports the hypothesis that the period of control for the medial consonant is superimposed onto underlying periods of control for the flanking vowels. Since vowel production involves (mainly) the tongue body and jaw, and most consonants involve the jaw as well, then during periods of coproduction the influences of the overlapping gestures must be blended at the level of the shared articulators.

Interarticulatory Coordination; Single Speech Gestures

In the task dynamic model, coordinative dynamics are posited at an abstract level of system description, and give rise to appropriately gesture-specific and contextually variable patterns at the level of articulatory motions. Since one of the major tasks for speech is to create and release constrictions in different local regions of the vocal tract, the abstract dynamics are defined in coordinates that represent the configurations of different constriction types, e.g., the bilabial constrictions used in producing /b/, /p/, or /m/, the alveolar constrictions used in producing /d/, /t/, or /n/, etc. Typically, each constriction type is associated with a pair of so-called *tract variable* coordinates, one that refers to the location of the constriction along the longitudinal axis of the vocal tract, and one that refers to the degree of constriction measured perpendicularly to the longitudinal axis in the midsagittal plane. For example, bilabial constrictions are defined according to the tract variables of lip aperture and lip protrusion (see Figure 2). Lip aperture defines the degree of bilabial constriction, and is defined by the vertical distance between the upper and lower lips; lip protrusion defines the location of bilabial constriction, and is defined by the horizontal distance between the (yoked) upper and lower lips and the upper and lower front teeth, respectively. Constrictions are restricted to two dimensions for practical purposes, due to the fact that the simulations use the articulatory geometry represented in the Haskins Laboratories software articulatory synthesizer (Rubin, Baer, & Mermelstein, 1981). This synthesizer is defined according to a midsagittal representation of the vocal tract, and converts a given articulatory configuration in this plane, first to a sagittal vocal tract outline, then to a three dimensional tube shape, and finally, with the addition of appropriate voice source information, to an acoustic waveform. As a working hypothesis, the tract-variable gestures in the model have been assigned the point attractor dynamics of damped, second order systems, analogous to those of damped mass-spring systems. Each gesture is assigned its own set of dynamic parameters: target or rest position, natural frequency, and damping factor. Gestures are active over discrete time intervals, e.g., over discrete periods of bilabial closing or opening, laryngeal abduction or adduction, tongue-tip raising or lowering, etc.

Just as each constriction type is associated with a set of tract variables, each tract variable is associated with a set of *model articulator* coordinates that comprises an articulatory subset for the tract variable. The model articulators are defined according to the articulatory degrees-of-freedom of the Haskins software synthesizer. Figure 2 shows the relation between tract variable and model articulator coordinates (see also Figure 2 in Browman & Goldstein, in press). The model articulators are controlled by transforming the tract-variable dynamical system into model articulator coordinates. This coordinate transformation creates a set of gesture-specific and articulatory posture-specific coupling functions among the articulators. These functions create a dynamical system at the articulatory level whose modal, cooperative behaviors allow them to flexibly and autonomously attain speech relevant goals. In other words, the tract-variable coordinates define a set of gestural modes for the model articulators (see also Coker, 1976 for a related treatment of vocal tract modes).

Tract variables		Model articulators
LP	lip protrusion	upper & lower lips
LA	lip aperture	upper & lower lips, jaw
TDCL	tongue dorsum constrict location	tongue body, jaw
TDCD	tongue dorsum constrict degree	tongue body, jaw
LTH	lower tooth height	jaw
TTCL	tongue tip constrict location	tongue tip,body, jaw
TTCD	tongue tip constrict degree	tongue tip,body, jaw
VEL	velic aperture	velum
GLO	glottal aperture	glottis

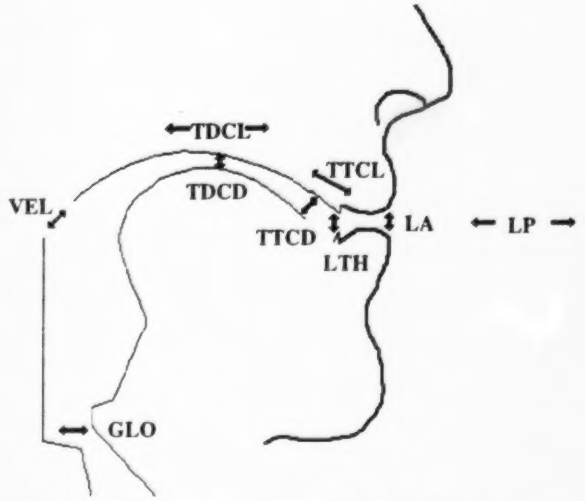


Figure 2. Top: Table showing the relationship between tract-variables and model articulators; Bottom: Schematic midsagittal vocal tract outline, with tract-variable degrees of freedom indicated by arrows. (from "The Task Dynamic Model in Speech Production" by E. Saltzman, 1991; reprinted by permission).

Significantly, articulatory movement trajectories unfold as implicit consequences of the tract-variable dynamics without reference to explicit trajectory plans or templates. Additionally, the model displays gesture-specific patterns of remote compensation to simulated mechanical perturbations delivered to the model articulators (Figure 3), that mirror the compensatory effects reported in the experimental literature (Figure 1). In particular, simulations were performed of perturbed and unperturbed bilabial closing gestures (Saltzman, 1986; Kelso, Saltzman & Tuller, 1986a, 1986b). When the simulated jaw was "frozen" in place during the closing gesture, the system achieved the same final degree of bilabial closure in both the perturbed and unperturbed cases, although with different final articulatory configurations. Furthermore, the lips compensated spontaneously and immediately to the jaw perturbation, in the sense that neither replanning or reparameterization was required in order to compensate. Rather, compensation was brought about through the automatic and rapid redistribution of activity over the entire articulatory subset in a gesture-specific manner. The interarticulatory processes of control and coordination were exactly the same during both perturbed and unperturbed simulated gestures (see Kelso, et al., 1986a, 1986b, and Saltzman, 1986, for the mathematical details underlying these simulations).

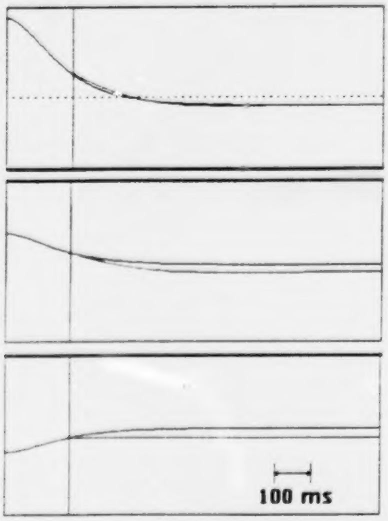


Figure 3. **Simulated tract variable and articulator trajectories for unperturbed (solid lines) and perturbed (dotted lines) bilabial closing gestures. Top panel: Lip aperture; Middle panel: Upper lip; Bottom panel: Jaw; Panels are aligned with reference to the perturbation onset (Solid vertical lines). Dashed horizontal line in top panel denotes zero lip aperture, with negative aperture signifying lip compression. (adapted from Figures 2, 3, & 4 in "Intentional contents, communicative context, and task dynamics: A reply to the commentators" by J. A. S. Kelso, E. L. Saltzman, & B. Tuller, 1986).**

Intergestural Coordination; Activation; Blending

How might gestures be combined to simulate speech sequences? In order to model the spatiotemporal orchestration of gestures evident in even the simplest utterances, a third coordinate system composed of gestural *activation* coordinates was defined. Each gesture in the model's repertoire is assigned its own activation coordinate, in addition to its set of tract-variables and model articulators. A given gesture's ongoing activation value defines the strength with which the gesture "attempts" to shape vocal tract movements at any given point in time according to its own phonetic goals (e.g., its tract variable target and natural frequency parameters). Thus, in its current formulation the task dynamic model of speech production is composed of two functionally distinct but interacting levels (see Figure 4). The *intergestural coordination* level is defined according to the set of gestural activation coordinates, and the *interarticulator coordination* level is defined according to both model articulator and tract-variable coordinates. The architectural relationships among these coordinates are shown in Figure 5.

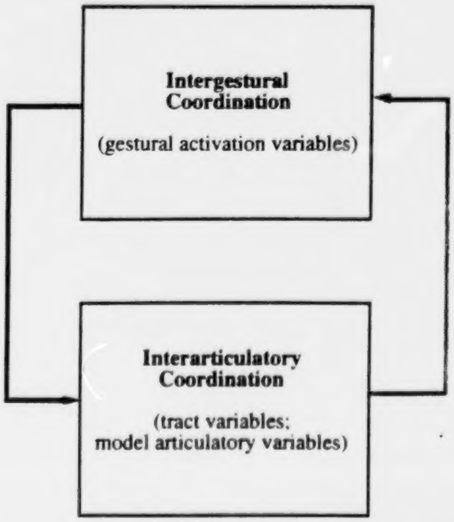

Intergestural Coordination
(gestural activation variables)
Interarticulatory Coordination
(tract variables; model articulatory variables)

Figure 4. **Schematic illustration of the two-level dynamical model for speech production, with associated coordinate systems indicated. The darker arrow from the intergestural to the interarticulator level denotes the feedforward flow of gestural activation. The lighter arrow indicates feedback of ongoing tract-variable and model articulator state information to the intergestural level. (from "A Dynamical Approach to Gestural Patterning in Speech Production" by E. L. Saltzman & K. G. Munhall, 1987; reprinted by permission).**

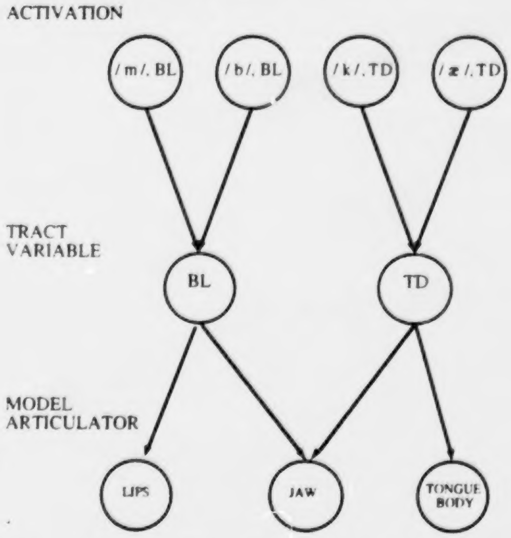


Figure 5. **Example of the "anatomical" relationships defined among model-articulator, tract-variable, and activation coordinate systems. BL and TD denote tract-variables associated with bilabial and tongue-dorsum constrictions, respectively. Gestures at the activation level are labeled in terms of both linguistic identity (e.g., /k/) and tract-variable affiliation (e.g., TD). (from "The Task Dynamic Model in Speech Production" by E. Saltzman, 1991; reprinted by permission).**

In current simulations, the gestural activation trajectories are defined for simplicity's sake as step functions of time, normalized from zero to one. Thus, outside a gesture's temporal interval of activation (i.e., when activation is zero), the gesture is inactive or "off" and has no influence on vocal tract activity. During its activation interval, when its activation value is one, the gesture is "on" and has maximal effect on the vocal tract. Viewed from this perspective, the problem of coordination among the gestures participating in a given utterance, e.g., for tongue-dorsum and bilabial gestures in a vowel-bilabial-vowel sequence, becomes that of specifying patterns of relative timing and cohesion among activation intervals for those gestures (see Saltzman & Munhall, 1989, for further details of the manner in which gestural activations influence vocal tract movements). Currently, intergestural relative timing patterns are specified by *gestural scores* that are generated explicitly either "by hand", or according to a linguistic gestural model that embodies the rules of Browman & Goldstein's *articulatory phonology* (e.g., Browman & Goldstein, 1986, 1991, in press). The manner in which gestural scores represent the relative timing patterns for an utterance's set of tract-variable gestures is shown in Figure 6 for the word "pub".

Using these methods, the task-dynamic model has been shown to reproduce many of the coproduction and intergestural blending effects found in the speech production literature. In the model, coproduction effects are generated as the articulatory and acoustic consequences of temporal overlap in gestural activations; blending occurs when there is spatial overlap of the gestures involved, i.e., when the gestures share model articulators in common. Blending would occur, for example, during coproduction of vowel (tongue and jaw) and bilabial (lips and jaw) gestures at the shared jaw articulator. The magnitude of coproduction effects is a function of the degree of spatial overlap of the gestures involved, i.e., the degree to which articulators are shared across gestures. Minimal interference occurs as long as the spatial overlap is incomplete.

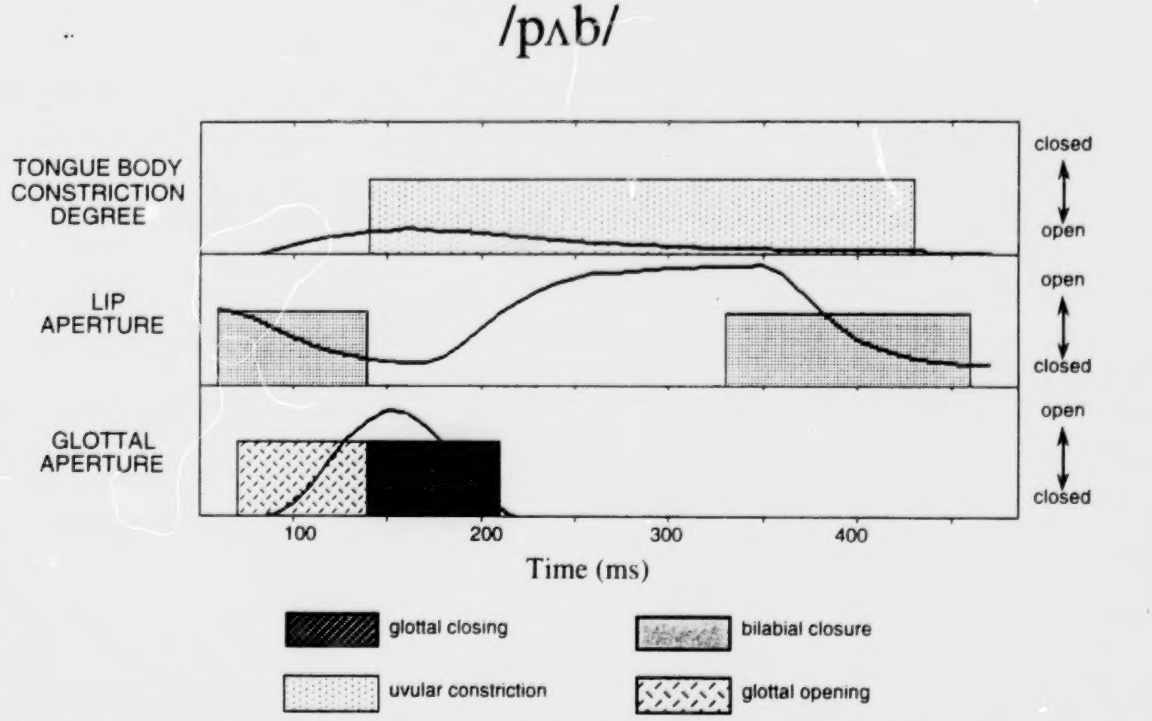


Figure 6. **Gestural score for the simulated sequence /pʌb/. Filled boxes denote intervals of gestural activation. Box heights are either 0 (no activation) or 1 (full activation). The waveform lines denote tract-variable trajectories produced during the simulation. (from "A Dynamical Approach to Gestural Patterning in Speech Production" by E. L. Saltzman & K. G. Munhall, 1987; reprinted by permission).**

This is the case when gestures are defined along distinct sets of tract-variables, and the gestures share none, or some but not all, articulators in common (see Figure 2). In this situation, the coproduced gestures can each attain their individual phonetic goals. Figure 7A illustrates the behavior of the model for two VCV sequences in which symmetric flanking vowels, /i/ and /æ/, vary across sequences, the medial consonant is the alveolar /d/ in both sequences, and the time courses of vowel and consonant activations are identical in both sequences. Vowels are produced using the tract-variables of tongue-dorsum constriction location and degree, and the associated jaw and tongue-body model articulators; the alveolar is produced using the tract-variables of tongue-tip constriction location and degree, and the associated jaw, tongue-body, and tongue-tip articulators. Thus, the vowel and consonant gestures share some but not all articulators in common. In this case, the alveolar's tongue-tip constriction goals are met identically in both sequences, although contextual differences in articulatory positions are evident, and are related to corresponding differences in the identities of the flanking vowels (for comparison, see the simulated tract shapes of isolated, steady-state productions of the vowels /i/ and /æ/, shown in Figure 7C).

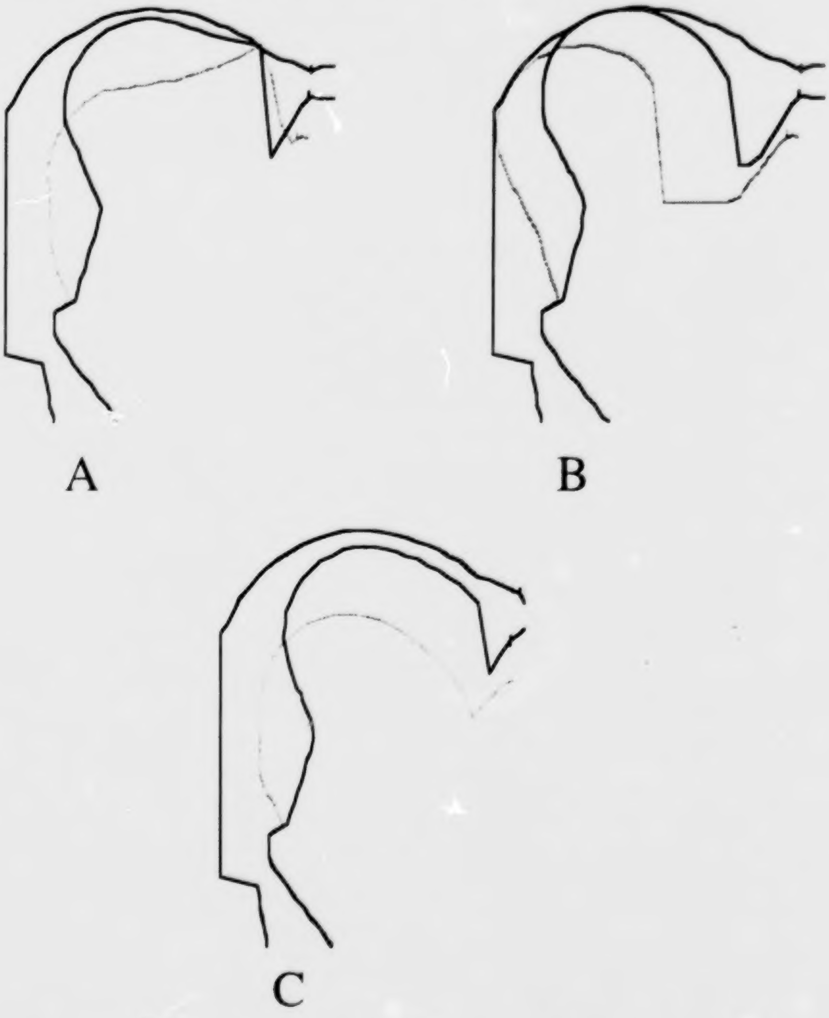


Figure 7. **Simulated vocal tract shapes. A. First contact of tongue-tip and upper tract wall during symmetric vowel-alveolar-vowel sequences; B. First contact of tongue-dorsum and upper tract wall during symmetric vowel-velar-vowel sequences; C. Corresponding steady-state vowel productions. (Dark lines denote /i/ tokens, light lines denote /æ/ tokens). (from "The Task Dynamic Model in Speech Production" by E. Saltzman, 1991; reprinted by permission).**

However, when coproduced gestures use the same sets of tract variables, all articulators are shared in common, and there is the potential for mutual interference in attaining competing phonetic goals. Figure 7B illustrates the behavior of the model for two VCV sequences that are identical to those shown in Figure 7A, except that the medial consonant is the velar /g/. In this situation, consonant and vowels are produced using the same tongue-dorsum tract variables and the same jaw and tongue-body model articulators. During periods of coproduction the gestures compete for control of tongue-dorsum motion, resulting in contextual variation even in the attainment of the constriction target for /g/. The velar's place of constriction is altered by the identity of the flanking vowels, although the degree of constriction is not. Importantly, the simulations displayed in Figures 7A & 7B mirror the patterns observed experimentally during actual VCV production (Öhman, 1967). Additionally, such processes of within-tract variable blending are consistent with data on experimentally-induced vowel production errors (Laver, 1980), in which blended vowel forms were produced that were intermediate between canonical forms.

Future Directions

In its current state, the task dynamic model offers a useful and promising account of movement patterns observed during unperturbed and mechanically perturbed speech sequences, and during periods of coproduction. Significantly, explicit trajectory planning is not required, and the model functions in exactly the same way during simulations of unperturbed, mechanically perturbed, and coproduced speech gestures. Additionally, the model provides a way to reconcile much of the apparent conflict between observations of surface articulatory and acoustic variability on the one hand, and the hypothesized existence of underlying, invariant linguistic units on the other hand. Invariant units are specified in the form of context-independent sets of gestural parameters (e.g., tract variable targets), and are associated with corresponding subsets of activation, tract-variable, and articulator coordinates. Variability emerges in the tract-variable and articulatory movement patterns, as a result of both the utterance-specific temporal interleaving of gestural activations provided by the gestural scores, and the accompanying dynamics of intergestural blending during coproduction.

One of the main drawbacks of the model from a dynamical perspective is that there are no dynamics intrinsic to the level of intergestural coordination that are comparable to the dynamics intrinsic to the interarticulatory level. The patterning of gestural activation trajectories is specified explicitly either "by hand" or by the rules embodied in the linguistic gestural model of Browman & Goldstein. Once a gestural score is specified, it remains fixed throughout a given simulation, defining a unidirectional, rigidly feedforward flow of control from the intergestural to interarticulator levels of the model. The gestural score acts, in essence, like the punched paper roll that drives the keys of a player piano. Experimental data suggest, however, that the situation is not this simple. For example, transient mechanical perturbations delivered to the speech articulators during repetitive speech sequences (Saltzman, 1992; Saltzman, Kay, Rubin, & Kinsella-Shaw, 1991), or to the limbs during unimanual rhythmic tasks (Kay, 1986; Kay, et al., 1991), can alter the underlying timing structure of the ongoing sequence and induce systematic shifts in the timing of subsequent movement elements. These data imply that activation patterns are not rigidly specified over a given sequence. Rather, such results suggest that activation trajectories evolve fluidly and flexibly over the course of an ongoing sequence governed by an intrinsic intergestural dynamics, and that this intergestural dynamical system functions as a sequence-specific timer or clock that is bidirectionally coupled to the interarticulator level.

Work is currently in progress (with colleagues John Hogden, Simon Levy, and Philip Rubin) to incorporate the dynamics of connectionist networks (e.g., Bailly, Laboissiere, & Schwartz, 1991; Grossberg, 1986; Jordan, 1986, 1990, in press; Kawato, 1989) at the intergestural level of the model, in order to shape activation trajectories intrinsically and to allow for adaptive online interactions with the interarticulatory level. In particular, we adopted the recurrent, sequential network architecture of Jordan (1986, 1990, in press). Each output node of the network represents a corresponding gestural activation coordinate. The values of these output nodes range continuously from zero to one, allowing each gesture's influence over the vocal tract to wax and wane in a smoothly graded fashion. Additionally, the ongoing tract-variable state will be fed back into the sequential net, providing an informational basis for the modulation of activation timing patterns by simulated perturbations delivered to the model articulator or tract-variable coordinates. Thus, rather than being explicitly and rigidly determined prior to the onset of the simulated utterance, the activation patterns will evolve during the utterance as implicit consequences of the dynamics of the entire multilevel (intergestural and interarticulatory) system.

4. SUMMARY AND CONCLUSIONS

The dynamical approach described in this chapter provides a powerful set of empirical and theoretical tools for investigating and understanding the coordination and control of skilled sensorimotor activities, ranging from simple one-joint rhythms to the complex patterns of speech production. The approach offers a unified and rigorous account of a movement's spatiotemporal form, stability of form, lawful warpings of form induced by scaling performance parameters, and the intuitive relation between underlying invariance and surface variability. Evidence was reviewed supporting the hypothesis that dynamical systems governing skilled sensorimotor behaviors are defined in abstract, low-dimensional, task-spaces that serve to create modal or cooperative patterns of activity in the generally higher-dimensional articulatory periphery. In this regard, the single and dual degree of freedom limb rhythms, considered in the *Dynamics* section of the chapter, can be viewed as tasks with relatively simple mappings between their respective task (or modal) coordinates and articulatory coordinates. Such tasks are rare in everyday life, however. Most real world activities (e.g., speech production, or the coordination of reaching and grasping for object retrieval and manipulation) involve tasks defined over effector systems with multiple articulatory degrees of freedom, and for which the mappings between task and articulatory coordinates are more complex.

The abstract nature of these coordinative dynamics was highlighted by the demonstration (Schmidt, et al., 1990) that entrainment between two limit cycle rhythms can occur when the component rhythms are performed by different actors, who are linked by visual information. These data suggest that the intent to coordinate one's actions with events in the external environment serves to create a linkage through which perceptual information, specific to the dynamics of these events, flows into the component task-spaces that control these actions. The result is a coupled, abstract, modal dynamical system that seamlessly spans actor and environment. It is tempting to speculate that this perspective applies quite generally across the spectrum of biological behaviors.

REFERENCES

Abbs, J. H., & Gracco, V. L. (1983). Sensorimotor actions in the control of multimovement speech gestures. *Trends in Neuroscience, 6,* 393-395.

Abraham, R., & Shaw, C. (1982). *Dynamics-The geometry of behavior. Part 1: Periodic behavior.* Santa Cruz, CA: Aerial Press.

Bailly, G., Laboissière, R., & Schwartz, J. L. (1991). Formant trajectories as audible gestures: An alternative for speech synthesis. *Journal of Phonetics, 19,* 9-23.

Baker, G. L. & Gollub, J. P. (1990). *Chaotic dynamics: An introduction.* New York: Cambridge University Press.

Baldissera, F., Cavallari, P., & Civaschi, P. (1982). Preferential coupling between voluntary movements of ipsilateral limbs. *Neuroscience Letters, 34,* 95-100.

Baldissera, F., Cavallari, P., Marini, G., & Tassone, G. 1991. Differential control of in-phase and anti-phase coupling of rhythmic movements of ipsilateral hand and foot. *Experimental Brain Research, 83,* 375-380.

Beek, P. J. 1989. Timing and phase-locking in cascade juggling. *Ecological Psychology, 1,* 55-96.

Bell-Berti, F., & Harris, K. S. 1981. A temporal model of speech production. *Phonetica, 38,* 9-20.

Bernstein, N. A. (1967). *The coordination and regulation of movements.* London: Pergamon Press. Reprinted in H. T. A. Whiting (Ed.), *Human motor actions: Bernstein reassessed.* New York: North-Holland. 1984.

Browman, C., & Goldstein, L. (1986). Towards an articulatory phonology. In C. Ewan, & J. Anderson (Eds.), *Phonology yearbook 3* (pp. 219-252). Cambridge: Cambridge University Press.

Browman, C. P., & Goldstein, L. (1991). Tiers in articulatory phonology, with some implications for casual speech. In J. Kingston & M. E. Beckman (Eds.), *Papers in Laboratory Phonology: I. Between the Grammar and the Physics of Speech.* (pp. 341-338). Cambridge, England: Cambridge University Press.

Browman, C. P., & Goldstein, L. (in press). Dynamics and articulatory phonology. In T. van Gelder & R. Port (Eds.), *Mind as motion.* MIT Press.

Coker, C. H. (1976). A model of articulatory dynamics and control. *Proceedings of the IEEE, 64,* 452-460.

de Jong, K. 1991. An articulatory study of consonant-induced vowel duration changes in English. *Phonetica, 48,* 1-17.

Edwards, J., Beckman, M. E., & Fletcher, J. (1991). The articulatory kinematics of final lengthening. *Journal of the Acoustical Society of America, 89,* 369-382.

Folkins, J. W., & Abbs, J. H. (1975). Lip and jaw motor control during speech: Responses to resistive loading of the jaw. *Journal of Speech and Hearing Research, 18,* 207-220.

Fowler, C. 1980. Coarticulation and theories of extrinsic timing. *Journal of Phonetics, 8,* 113-133.

Fowler, C. A., & Saltzman, E. 1993. Coordination and coarticulation in speech production. *Language and Speech, 36,* 171-195.

Grossberg, S.1986. The adaptive self-organization of serial order in behavior: Speech, language, and motor control. In E. C. Schwab & H. C. Nusbaum (Eds.), *Pattern recognition by humans and machines (Vol. 1).* New York: Academic Press.

Haken, H., Kelso, J. A. S., & Bunz, H. (1985). A theoretical model of phase transitions in human hand movements. *Biological Cybernetics, 51,* 347-356.

Harris, K. S. (1984). Coarticulation as a component of articulatory descriptions. In R. G. Daniloff (Ed.), *Articulation assessment and treatment issues* (pp. 147-167). San Diego, CA: College Hill Press.

Hawkins, S. (1992). An introduction to task dynamics. In G. J. Docherty & D. R. Ladd (Eds.), *Papers in laboratory phonology. II. Gesture, segment, and prosody* (pp. 9-25). Cambridge: Cambridge University Press.

BEST COPY AVAILABLE

Hollerbach, J. M. (1982). Computers, brains, and the control of movement. *Trends in Neurosciences, 5*, 189-192.

Jordan, M. I. 1986. *Serial order in behavior: A parallel distributed processing approach* (Tech. Rep. No. 8604). San Diego: University of California, Institute for Cognitive Science.

Jordan, M. I. (1990). Motor learning and the degrees of freedom problem. In M. Jeannerod (Ed.), *Attention and performance XIII* (pp. 796-836). Hillsdale, NJ: Erlbaum.

Jordan, M. I. (in press). Serial order: A parallel distributed processing approach. In J. L. Elman & D. E. Rumelhart (Eds.), *Advances in connectionist theory: Speech*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Jordan, M. I., & Rosenbaum, D. A. (1989). Action. In M. I. Posner (Ed.), *Foundations of cognitive science* (pp. 727-767). Cambridge, MA: MIT Press.

Kawato, M. (1989). Motor theory of speech perception revisited from minimum torque-change neural network model. *Proceedings of the 8th Symposium on Future Electron Devices, October 30-31, Tokyo, Japan* (pp. 141-150).

Kay, B. A. (1986). *Dynamic modeling of rhythmic limb movements: Converging on a description of the component oscillators*. Unpublished doctoral dissertation, Department of Psychology, University of Connecticut, Storrs.

Kay, B. A., Kelso, J. A. S., Saltzman, E. L., & Schöner, G. (1987). Space-time behavior of single and bimanual rhythmical movements: Data and limit cycle model. *Journal of Experimental Psychology: Human Perception and Performance, 13*, 178-192.

Kay, B. A., Saltzman, E. L., & Kelso, J. A. S. (1991). Steady-state and perturbed rhythmical movements: A dynamical analysis. *Journal of Experimental Psychology: Human Perception and Performance, 17*, 183-197.

Keating, P. A. (1985). CV phonology, experimental phonetics, and coarticulation. *UCLA Working Papers in Phonetics, 62*, 1-13.

Kelso, J. A. S. (1984). Phase transitions and critical behavior in human bimanual coordination. *American Journal of Physiology: Regulatory, Integrative and Comparative Physiology, 15*, R1000-R1004.

Kelso, J. A. S., Buchanan, J. J., & Wallace, S. A. (1991). Order parameters for the neural organization of single, multijoint limb movement patterns. *Experimental Brain Research*, 85:432-444.

Kelso, J. A. S., Delcolle, J. D., & Schöner, G. S. (1990). Action-perception as a pattern formation process. In M. Jeannerod (Ed.), *Attention and performance XIII* (pp. 139-169). Hillsdale, NJ: Erlbaum.

Kelso, J. A. S., & Jeka, J. J. (1992). Symmetry breaking dynamics of human multilimb coordination. *Journal of Experimental Psychology: Human Perception and Performance, 18*, 645-668.

Kelso, J. A. S., Saltzman, E. L., & Tuller, B. (1986a). The dynamical perspective on speech production: Data and theory. *Journal of Phonetics, 14*, 29-60.

Kelso, J. A. S., Saltzman, E. L., & Tuller, B. (1986b). Intentional contents, communicative context, and task dynamics: A reply to the commentators. *Journal of Phonetics, 14*, 171-196.

Kelso, J. A. S., Tuller, B., Vatikiotis-Bateson, E., & Fowler, C. A. (1984). Functionally specific articulatory cooperation following jaw perturbations during speech: Evidence for coordinative structures. *Journal of Experimental Psychology: Human Perception and Performance, 10*, 812-832.

Kelso, J. A. S., Vatikiotis-Bateson, E., Saltzman, E. L., & Kay, B. A. (1985). A qualitative dynamic analysis of reiterant speech production: Phase portraits, kinematics, and dynamic modeling. *Journal of the Acoustical Society of America, 77*, 266-280.

Kent, R. D., & Minifie, F. D. (1977). Coarticulation in recent speech production models. *Journal of Phonetics, 5*, 115-133.

Kugler, P. N., & Turvey, M. T. (1987). *Information, natural law, and the self-assembly of rhythmic movement*. Hillside, NJ: Erlbaum Associates.

Laver, J. (1980). Slips of the tongue as neuromuscular evidence for a model of speech production. In H. W. Dechert & M. Raupach (Eds.), *Temporal variables in speech: Studies in honour of Frieda Goldman-Eisler*. The Hague: Mouton.

MacKay, D. G., & Soderberg, G. A. (1971). Homologous intrusions: An analogue of linguistic blends. *Perceptual and Motor Skills, 32*, 645-646.

MacKenzie, C. L., & Patla, A. E. (1983). Breakdown in rapid bimanual finger tapping as a function of orientation and phasing. *Society for Neuroscience Abstracts, 9* (2).

Mattingly, I. (1990). The global character of phonetic gestures. *Journal of Phonetics, 18*, 445-452.

Norton, A. in press. Dynamics: An introduction. In T. van Gelder & R. Port (Eds.), *Mind as motion*. MIT Press.

Öhman, S. E. G. (1966). Coarticulation in VCV utterances: Spectrographic measurements. *Journal of the Acoustical Society of America, 39*, 151-168.

Öhman, S. E. G. (1967). Numerical model of coarticulation. *Journal of the Acoustical Society of America, 41*, 310-320.

Perkell, J. S. (1969). *Physiology of speech production: Results and implications of a quantitative cineradiographic study*. Cambridge, MA: MIT Press.

Perkell, J. S. (1991). Models, theory, and data in speech production. *Proceedings of the XIIth International Congress of Phonetic Sciences, volume 1*. Aix-en-Provence, France: Universte de Provence Service des Publications.

Rosenblum, L. D., & Turvey, M. T. (1988). Maintenance tendency in coordinated rhythmic movements: Relative fluctuations and phase. *Neuroscience, 27*, 289-300.

Rubin, P. E., Baer, T., & Mermelstein, P. (1981). An articulatory synthesizer for perceptual research. *Journal of the Acoustical Society of America, 70*, 321-328.

Saltzman, E. L. (1979). Levels of sensorimotor representation. *Journal of Mathematical Psychology, 20*, 91-163.

Saltzman, E. (1986). Task dynamic coordination of the speech articulators: A preliminary model. *Experimental Brain Research*, Ser. 15, 129-144.

Saltzman, E. (1991). The task dynamic model in speech production. In H. F. M. Peters, W. Hulstijn, & C. W. Starkweather (Eds.), *Speech motor control and stuttering*. (pp. 37-52) Amsterdam: Excerpta Medica.

Saltzman, E. L. (1992). Biomechanical and haptic factors in the temporal patterning of limb and speech activity. *Human Movement Science, 11*, 239-251.

Saltzman, E., Kay, B., Rubin, P., & Kinsella-Shaw, J. (1991). Dynamics of intergestural timing. *Perilus XIV*, Institute of Linguistics, University of Stockholm, Stockholm, Sweden. (pp. 47-56).

Saltzman, E. L., & Kelso, J. A. S. (1987). Skilled actions: A task dynamic approach. *Psychological Review, 94*, 84-106.

Saltzman, E. L. & Munhall, K. G. 1989. A dynamical approach to gestural patterning in speech production. *Ecological Psychology, 1*, 333-382.

Schmidt, R. C., Carello, C., & Turvey, M. T. (1990). Phase transitions and critical fluctuations in the visual coordination of rhythmic movements between people. *Journal of Experimental Psychology: Human Perception and Performance, 16*, 227-247.

Schneider, K., Zernicke, R. F., Schmidt, R. A., & Hart, T. J. (1989). Changes in limb dynamics during practice of rapid arm movements. *Journal of Biomechanics, 22*, 805-817.

Scholz, J. P., & Kelso, J. A. S. (1989). A quantitative approach to understanding the formation and change of coordinated movement patterns. *Journal of Motor Behavior, 21*, 122-144.

Schöner, G., Haken, H., & Kelso, J. A. S. (1986). Stochastic theory of phase transitions in human hand movement. *Biological Cybernetics, 53,* 1-11.

Schöner, G., & Kelso, J. A. S. (1988). Dynamic pattern generation in behavioral and neural systems. *Science,* 239, 1513-1520.

Shaiman, S. (1989). Kinematic and electromyographic responses to perturbation of the jaw. *Journal of the Acoustical Society of America,* 86, 78-88.

Soechting, J. F. (1982). Does position sense at the elbow joint reflect a sense of elbow joint angle or one of limb orientation? *Brain Research, 248,* 392-395.

Sternad, D., Turvey, M. T., & Schmidt, R. C. (1992). Average phase difference theory and 1:1 phase entrainment in interlimb coordination. *Biological Cybernetics, 67,* 223-231.

Sussman, H. M., MacNeilage, P. F., & Hanson, R. J. (1973). Labial and mandibular dynamics during the production of bilabial consonants: Preliminary observations. *Journal of Speech and Hearing Research, 16,* 397-420.

Thompson, J. M. T., & Stewart, H. B. (1986). *Nonlinear dynamics and chaos: Geometrical methods for engineers and scientists.* New York: Wiley.

Turvey, M. T. (1990). Coordination. *American Psychologist, 45,* 938-953.

Turvey, M. T., & Carello, C. (in press). Some dynamical themes in perception and action. In T. van Gelder & R. Port (Eds.), *Mind as motion.* MIT Press.

van Riel, M.-J., Beek, P. J., & van Wieringen, P. C. W. (1991). Phase transitions in rhythmic arm movements under different stimulus-response configurations. In P. J. Beek, R. J. Bootsma, & P. C. W. van Wieringen (Eds.), *Studies in perception and action* (pp. 234-238). Amsterdam: Rodopi.

Vatikiotis-Bateson, E. (1988). *Linguistic structure and articulatory dynamics.* Ph.D. Dissertation, Indiana University, Bloomington, IN (distributed by the Indiana University Linguistics Club, Bloomington, IN).

Woods, B. T., & Teuber, H.-L. (1978). Mirror movements after childhood hemiparesis. *Neurology, 28,* 1152-1158.

GUIDELINES FOR FURTHER READING

The Russian motor physiologist, N. A. Bernstein (1967/1984) produced a classic body of empirical and theoretical work that anticipated and inspired many of today's developments in movement science. It's still a great read. Turvey (1990) reviews and extends this perspective in a broad overview of issues faced in studying the dynamics of coordination, carrying the reader on a tour from Bernstein to the current state of the art. Readers interested in more detailed accounts of various recent trends in the field should consult Jordan (1990; connectionist perspective on dynamics and coordinate systems in skilled actions), Saltzman & Munhall (1989; task dynamics and speech production), and Schöner & Kelso (1988; an overview of the "synergetics" approach to self-organizing systems, in the context of sensorimotor behaviors).

Bernstein, N. A. 1967. *The coordination and regulation of movements.* London: Pergamon Press. Reprinted in H. T. A. Whiting, ed. 1984. *Human motor actions: Bernstein reassessed.* New York: North-Holland.

Jordan, M. I. 1990. Motor learning and the degrees of freedom problem. In M. Jeannerod (Ed.), *Attention and performance XIII* (pp. 796-836). Hillsdale, NJ: Erlbaum.

Saltzman, E. L. & Munhall, K. G. 1989. A dynamical approach to gestural patterning in speech production. *Ecological Psychology, 1,* 333-382.

Schöner, G., & Kelso, J. A. S. 1988. Dynamic pattern generation in behavioral and neural systems. *Science, 239,* 1513-1520.

Turvey, M. T. 1990. Coordination. *American Psychologist, 45,* 938-953.

FOOTNOTES

[*] To appear in T. van Gelder & R. Port (Eds.), *Mind as motion.* MIT Press.

[†] Also Center for the Ecological Study of Perception and Action, University of Connecticut, Storrs.

[1] Similar results on rhythms produced at the elbow and wrist joints of the same arm were presented by Kelso, Buchanan, and Wallace (1991), when the forearm was either pronated or supinated across experimental conditions. Again, the easiest combinations to perform were those in which the motions of the hand and forearm were spatially inphase, regardless of the relative anatomical phasing between hand and forearm muscle groups. Furthermore, in trials involving experimentally demanded increases or decreases of coupled oscillation frequency, phase transitions were observed from the spatially antiphase to spatially inphase patterns in both pronation and supination conditions. Relatedly, MacKenzie and Patla (1983) induced phase transitions in bimanual finger rhythms by increasing cycling frequency within trials, and showed that the transitions were affected systematically by the relative orientation of the fingers' spatial planes of motion.

The primacy of abstract spatial coordinates over anatomical or biomechanical coordinates has also been demonstrated for discrete targeting tasks. For example, Soechting (1982) reported evidence from a pointing task involving the elbow joint suggesting that the controlled variable for this task is not anatomical joint angle *per se,* but rather the orientation angle of the forearm in body-referenced or environment-referenced coordinates.

Haskins Laboratories Status Report on Speech Research
1993, SR-115/116, 17-32

Speech Motor Coordination and Control: Evidence From Lip, Jaw, and Laryngeal Movements*

Vincent L. Gracco and Anders Löfqvist

The movements of the lower lip, jaw and larynx during speech were examined for two different speech actions involving oral closing for /p/ and oral constriction for /f/. The initial analysis focused on the manner in which the different speech articulators were coordinated to achieve sound production. It was found that the lip, jaw, and laryngeal movements were highly constrained in their relative timing apparently to facilitate their coordination. Differences were noted in the degree to which speech articulator timing covaried dependent on the functional characteristics of the action. Movements associated with coordinating multiple articulators for a single sound were more highly constrained in their relative timing than were movements associated with sequencing of individual sounds. The kinematic patterns for the different articulators were found to vary in a number of systematic ways depending on the identity of the sound being produced, the phonetic context surrounding the target sound, and whether one versus two consonants were produced in sequence. The results are consistent with an underlying organization reflecting the construct of the phoneme. It is suggested that vocal tract actions for the sounds of the language are stored in memory as motor programs and sequenced together into larger meaningful units during speaking. Speech articulator motion for the different vowel sounds was found to be influenced by the identity of the following consonant suggesting that speech movements are modified in chunks larger than the individual phonetic segments. It appears that speech production is a hierarchical process with multiple levels of organization transforming cognitive intent into coherent and perceptually identifiable sound sequences.

INTRODUCTION

As a highly developed skilled motor behavior, speech production provides a rich environment for observing the functional synergies and coordinative principles that underlie a uniquely human behavior. Like most motor behaviors, speaking requires the interaction of multiple effectors (speech articulators) into larger functional aggregates. These articulatory aggregates are the framework for speech motor control and their activation is associated with sound production. As such, the ultimate goal of speech movement coordination is generally known. One issue of interest in speech motor control as well as motor control in general is the manner in which the nervous system controls the multiple degrees of movement freedom (Bernstein, 1967). It is generally accepted that the nervous system employs simplifying strategies to reduce the potentially independent variables (motor units, muscles, joints) in most motor behaviors to a controllable number (Gracco, 1988; Lacquaniti & Soechting, 1982; MacKenzie, 1992; MacPherson, 1988 a & b; Soechting & Lacquaniti, 1989; Turvey, 1977). Recently, analysis of the relative timing of the lips and jaw suggest that the multiple articulators are interdependently modulated such that timing variation in one articulator is accompanied by proportional changes in the timing of all the active articulators (Gracco, 1988; Gracco, 1994; Gracco & Abbs, 1986). Rather than considering each articulator as independently controlled it has been suggested that speech articulators are functionally constrained. That is, rather than explicitly controlling the timing of the different neuromuscular ele-

---

This research was supported by NIH grants DC-00121, DC-00595, and DC-00865 from the National Institute on Deafness and Other Communication Disorders, and by Esprit-BR Project 6975-Speech Maps.

ments involved in the production of a particular sound, the nervous system controls the coordinative requirements of all the active effectors as a unit (Gracco, 1990; 1991)

To date the most direct evidence for constraining speech movement timing has come from a relatively simple articulatory event, oral closing (Gracco, 1994; Gracco, 1988; Gracco & Abbs, 1986). Oral closing for bilabial sounds such as /p/, /b/, or /m/, simply involves the approximation of the two lips momentarily. It is not clear how general such coordinative interactions are among different speech articulators and whether such interactions change for different speech sounds. As noted above speech production is dependent on the actions of articulators other than the lips and jaw. One speech articulator that is also involved in many of the sounds of English is the larynx. The larynx is a time-varying valve involved in the initiation and arrest of vocal fold vibration for various vowel and consonant sounds. For voiceless consonant sounds, such as /p/, /t/, /k/, /f/, /s/, /sh/, the larynx must open in conjunction with the raising of tongue or lips to create an occlusion (/p/, /t/, or /k/) or constriction (/f/, /s/, /sh/) generating the necessary aerodynamic conditions within the vocal tract. For vowels following a voiceless consonant, the vocal folds, in conjunction with the jaw and tongue, approximate to provide a vibrating sound source. For each of these situations, voiceless consonants and vowels, the laryngeal action must be integrated with the movements of other speech articulators. Examination of the timing relations among the component articulators should provide insight into the speech movement coordination process. One focus of the present investigation is to examine the manner and degree to which the lips and jaw are coupled in their timing to the larynx for the production of vowel and voiceless consonant sounds. By examining the relative timing among the lips, jaw and larynx it should be possible to evaluate the degree and character of the temporal coupling associated with the different sound categories (consonant versus vowels).

From previous investigations it is also clear that the principles of speech movement coordination are not rigidly specified and vary at least according to movement direction. For example, lip and jaw motion for oral closing is tightly coupled and the timing of each articulator demonstrates significant covariation (Gracco & Abbs, 1986; Gracco, 1988). For oral opening, however, these articulators do not display the same degree of temporal coupling (Gracco, 1988; Gracco, 1994). Rather, for oral opening associated with a vowel sound, the timing constraint among the lips and jaw is apparently relaxed compared to their timing during oral closing. One possibility is that oral opening, generally associated with vowel production, and oral closing, generally associated with consonant production, are two distinct classes of speech motor actions with different principles underlying their coordination and control. Moving toward a vowel target, for example, may not require the same degree of temporal coupling among the contributing articulators as moving toward certain consonant targets (Gracco, 1994). It may not be surprising, then, that the lips and jaw are not as tightly coupled in their timing for oral opening as for oral closing. However, as with previous investigations, the context in which such observations have been made have been limited. The present investigation will focus on a larger phonetic context than has been previously examined.

A final focus of the present investigation of some theoretical importance for speech motor control is determining the characteristics of the underlying neural representation. While it is generally agreed that speech motor output is dependent on some underlying neural representations (production units) the form of such representations have yet to be determined. One possibility is that the units for speech are motor programs uniquely specified for the individual sounds (phonemes) of the language (Gracco, 1990; 1991). This conceptualization would require a finite number of motor programs (one per phoneme) that would be activated and sequenced into larger aggregates associated with syllables, words, phrases to allow meaningful communication. An alternative conceptualization involves a set of fundamental articulatory actions or features that are assembled and coordinated according to the phonetic context of the message (Kelso, 1986). One way to distinguish between these two alternatives is to examine the changes with context across articulators. Contextual variations influencing more than a single articulator might suggest that the entire vocal tract is being manipulated rather than the action of a single articulator. The difference between these two alternatives relates to the size of the fundamental units for speech production (phonetic segments versus articulatory gestures) and the level at which control is exerted (single or multiple articulators). Through a detailed examination of the movement differences associated with different sounds in sequence it will be possible to identify the specific kinematic adjustments that differentiate sounds and provide an objective method of characterizing speech articulator actions.

Materials and Methods

Three adult males (aged 40-48 years) served as subjects for the present investigation. Articulatory motion of the upper lip, lower lip, and jaw in the horizontal and vertical dimensions and changes in glottal area (or aperture) were obtained. Movements of the lips and jaw were transduced optoelectronically using small light emitting diodes (LED's) placed midsagittally on the vermilion border of the upper and lower lips. Changes in the positions of the LED's were sensed by a planar diode located in the focal plane of a camera mounted on a tripod and placed approximately 25 inches from the subject. For jaw motion, a custom fitted splint was constructed for each subject which fit snugly around the lower molars on one side. A piece of stainless steel wire was molded into the splint and bent to exit the corner of the mouth with minimal obstruction to the subjects articulation. The wire was bent to the midsagittal plane and an LED was placed on the extension of the jaw splint close to the chin allowing direct transduction of jaw motion. Glottal aperture was obtained using transillumination of the larynx. A flexible endoscope with a d.c. light source was passed through the nose and suspended in the oropharynx. The endoscope provided a light source that was registered at a sensor secured to the neck and placed external and inferior to the thyroid cartilage. The luminance registered at the sensor has been shown to vary as a function of changes in glottal area associated with opening and closing the glottis for voiceless sounds (Baer, Löfqvist, & McGarr, 1983; Löfqvist & Yoshioka, 1980). Figure 1 is a schematic representing the experimental set-up. Lip, jaw, and glottal signals were sampled at 500 Hz (12 bit resolution) and subsequently smoothed (42 point triangular window) and numerically differentiated (central difference) in software.

Subjects repeated one of seven words in the carrier phrase "It's a _______ again" at a comfortable speaking rate and loudness. The words used contained one of four vowels in combination with either the voiceless consonants /p/ and /f/ or the consonant sequence /ft/. The words included:

1) sapapple
2) supper
3) suffer
4) safe
5) safety
6) sipping
7) sifting

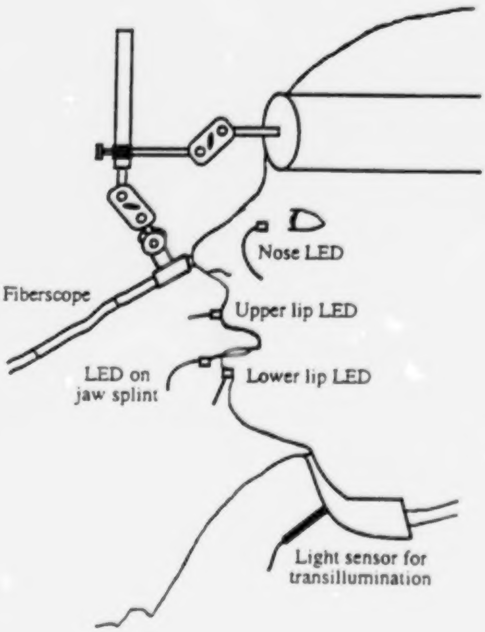


Figure 1. A line drawing of the experimental setup. A fiberscope, providing a d.c. light source, was passed through the nose and suspended in the pharynx. The light passed through the glottal opening in the larynx and the luminance was sensed from a sensor placed around the neck. The degree of luminance changed as a function of the glottal opening for the voiceless consonant sounds and was recorded as an analog voltage. Light emitting diodes (LED's) were placed on the bridge of the nose, the upper lip, lower lip, and on a jaw splint that exited from the mouth and provided signals corresponding to the motion of the respective articulators in the horizontal and vertical dimensions (see text for details). The LED's were pulsed and the light emitted was sensed at a planar diode located in the focal plane of a camera mounted on a tripod.

For "supper" and "suffer" the same vowel was used with a different following consonant; : "safe" and "safety" differ in the presence of the consonant sequence (/ft/); "sipping" and "sifting" differ by the consonant sequence and the identity of the voiceless consonant (/p/ versus /f/). The words were repeated in blocks of ten and each block was repeated four times. For subject ES, a number of repetitions were discarded because of poor transillumination signal quality due to the tongue obscuring part of the d.c. light source. The number of repetitions for each word per subject was: S:VG; 40, 40, 39, 40, 39, 40, 39; S:AL; 40, 40, 40, 37, 40, 40, 40; S:ES; 32, 32, 37, 23, 39, 39, 39 for words 1-7 respectively.

Presented in Figure 2 are the signals recorded and the measurement points identified. To evaluate articulator coordination two temporal intervals were examined in detail. These include the temporal relationship between (1) jaw lowering and glottal closing for the vowel following the initial voiceless consonant /s/, and, (2) glottal opening and lower lip raising for the occlusion (/p/) or constriction (/f/). In all cases, the relative timing of the articulatory events were based on the time of peak velocity and referenced to the peak glottal opening associated with the /ts/ in "It's" in the carrier phrase. Because most of the motion of the lip and jaw was confined to the vertical plane (with respect to gravity), the kinematic measures will focus on this single dimension. Movements examined included the jaw lowering displacement and velocity for the different vowels, the lip raising displacement and velocity for the consonants, and the glottal aperture velocity for the opening and closing phases for the different consonants.

Results

Speech movement coordination—Relative timing. In the present investigation the relative timing of the lip and jaw were examined with respect to the action of the larynx. All the words examined began with /s/ which requires a stable and high jaw position (relative to the maxilla) for the tongue articulation. In addition, /s/ is a voiceless consonant requiring larynx abduction or glottal opening. For the different sound sequences the jaw is then lowered from its relatively high position and the larynx is closed to allow phonation for the different vowels. In the present context this vowel related action was then followed by lower lip raising and glottal opening to produce the subsequent voiceless consonants.

The initial comparison focused on the timing of the jaw lowering and the larynx closing action associated with opening the oral aperture for the different vowel sounds. As mentioned above, all times are relative to the peak glottal opening for the "its" in the carrier phrase and all timing measures reflect the occurrence of peak velocity associated with the respective lip, jaw, or laryngeal actions. The left portion of Figure 3 presents scatterplots of the time of the jaw lowering peak velocity and the time of the peak glottal closing velocity for the different vowels for the three subjects.

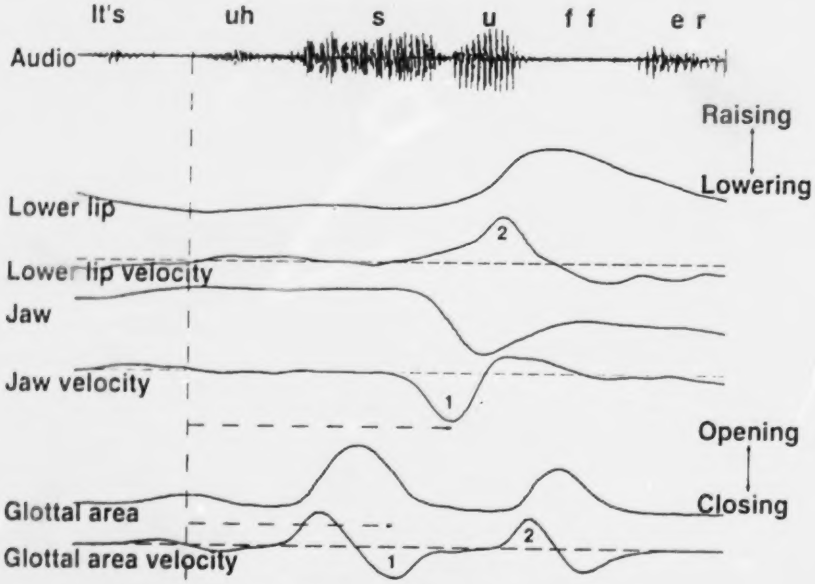


Figure 2. A schematic of the signals recorded for a single token of the phrase "It's a suffer" and the measurement points used in the present investigation. The signals from top to bottom include the acoustic signal recorded with a microphone, the vertical lower lip movement, the lower lip velocity, the vertical jaw movement, the jaw velocity, and the glottal area (aperture) and the change in glottal area (velocity). The dotted line indicates the midpoint of the glottal opening for "It's" and is used as the line-up point for all the timing measures (see text for details). The horizontal dotted lines illustrate one of the timing measures (1); the time of peak glottal closing and the time of peak jaw lowering for the vowel sound. In addition to the timing measures, the jaw lowering displacement and associated peak velocity and the lower lip raising and associated peak velocity and the peak glottal opening and closing velocities were also obtained.

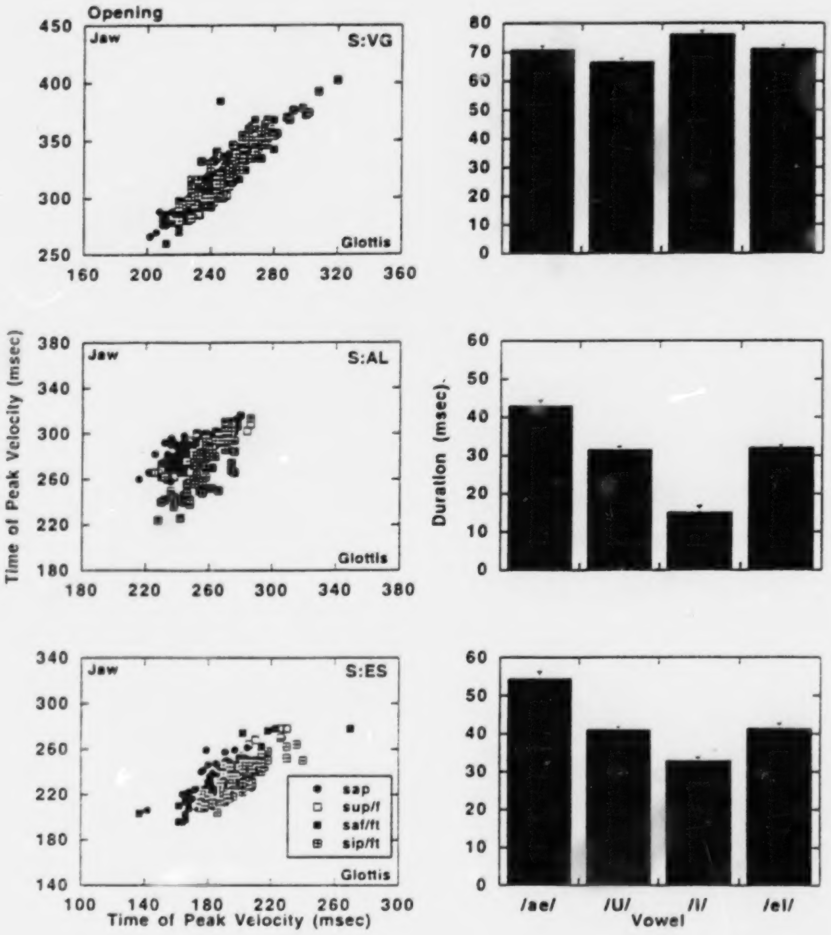


Figure 3. On the left, scatterplots of the time the jaw lowering (opening) peak closing velocity (in msec) as a function of the glottal peak closing velocity for the four different vowels (/ae/, /U/, /I/, and /ei/) for the three subjects. On the right are the mean differences between the time of the glottal closing velocity and jaw lowering velocity. The positive difference indicates that the time of glottal closing peak velocity always preceded the time of jaw lowering peak velocity. Error bars indicate one standard error.

The data have been grouped according to the different vowels following the /s/ sound. The vowel /U/ refers to "supper" and "suffer," /I/ refers to "sipping" and "sifting," /eI/ refers to "safe" and "safety," and /ae/ refers to "sapapple." As shown in the figure, there is a tendency for the timing of the jaw lowering to covary with the timing of the glottal closing. The correlation coefficients for the different vowels and subjects are presented in Table 1.

All correlations were significant ($p < .01$) although the magnitude of the relations varied quite a bit within and across the three subjects. For all subjects, the glottal closing peak velocity occurred in advance of the jaw lowering peak velocity.

Table 1. *Correlation of the time of the glottal closing velocity following /s/ with the time of peak velocity for the jaw lowering movement for the vowel for the three subjects.*

Subject	/ae/	/U/	/I/	/eI/
VG	.930	.699	.899	.930
AL	.549	.800	.886	.743
ES	.705	.893	.818	.864

This can be seen in the mean interval between the glottal closing peak velocity and the jaw lowering peak velocity presented in the right side of Figure 3. The positive value for each vowel indicates that the glottal adjustment is initiated prior to the jaw adjustment associated with the tongue action. For two of the three subjects the same trend was noted with the largest interval associated with the vowel /ae/ and the smallest interval associated with the vowel /I/. Interestingly, for these two subjects the intervals were positively correlated with the magnitude of the jaw lowering peak velocity (see below).

In contrast to the opening action, the closing action of the lips and larynx for the different consonant sounds was found to be highly correlated in relative timing. Correlation coefficients for the timing of lip raising and glottal opening are presented in Table 2.

In comparison to the correlations presented in Table 1, the correlations for the closing action were higher for all subjects with coefficients ranging from r = .93 to r = .99. Presented in the left portion of Figure 4 are scatterplots of the time of lip raising and glottal opening peak velocity for the three subjects. With few exceptions, the timing relations are similar across contexts.

Table 2. *Correlation of the time of peak lower lip raising velocity with the time of the glottal peak opening velocity for the consonant for the three subjects.*

Subject	/p/	/f/	/ft/
VG	.956	.941	.986
AL	.967	.974	.966
ES	.970	.982	.929

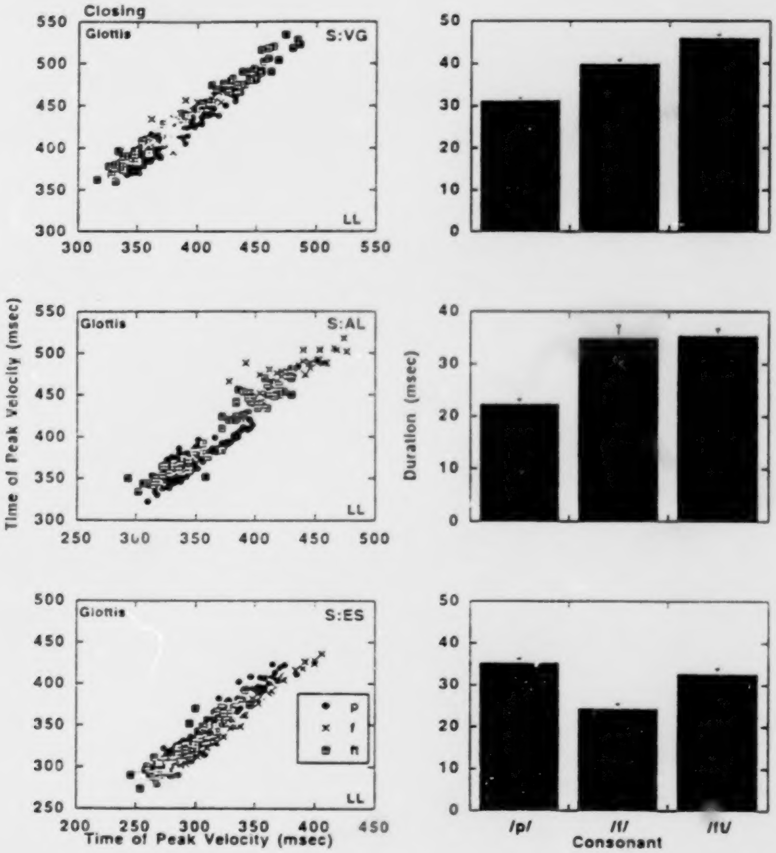


Figure 4. On the left, scatterplots of the time the lower lip raising (closing) peak velocity (in msec) as a function of the glottal opening peak velocity for the different consonants (/p/, /f/, and /ft/) for the three subjects. On the right are the mean differences between the time of the glottal opening peak velocity and lower lip raising peak velocity. The positive difference indicates that the time of lower lip raising peak velocity always preceded the time of glottal opening peak velocity. Error bars indicate one standard error.

Presented in the right portion of the figure are the mean intervals between the time of lip raising velocity and the time of glottal opening velocity for the three different consonant contexts. Similar to the opening sequence, the lower lip peak velocity always preceded the glottal opening peak velocity. Similar to the oral opening results, the rank order of the intervals was not consistent across the different subjects. Since the relation between lower lip raising and peak glottal opening actions may be an important variable associated with the different consonants, it was also of interest to determine whether these two events demonstrated systematic consonant-related changes. To address this issue the difference between the time of peak glottal opening and the time of peak displacement for lip raising was obtained. The results for the three subjects are presented in Figure 5. The positive values indicate that the glottal peak opening occurred after the peak raising displacement of the lower lip while the negative value for S:ES for /f/ indicates that the order was reversed. While the differences across the consonant conditions were statistically different, there was no consistent trend across the three subjects. However, it appears that for two subjects the interval for /f/ is smaller than for /p/ and the interval for /ft/ in longer than either of the single consonants.

Movement Adjustments

Oral opening. The results suggest that the lip, jaw and laryngeal movements are coupled in their timing and that the degree of coupling is greater for oral closing than for oral opening. In order to evaluate the manner in which these actions differ kinematically, the movements of the jaw, lip and larynx were examined in detail. As mentioned previously, all utterances examined were initiated from the same initial conditions. The different vowel sounds resulted in a range of jaw opening displacements and corresponding velocities. The first analysis focused on the relationship between jaw lowering displacement and velocity. As shown on the right side of Figure 6, the correlation of velocity and displacement is quite strong for all subjects. With the exception of the one cluster of data points for subject AL, each subjects' velocity/displacement relationship can be described by a single function. The left side of the figure presents the average jaw lowering displacement for the different vowels. It can be seen that jaw displacement varied in a systematic way for the different vowel sounds (see also Macchi, 1988; Oshima & Gracco, 1992). Of the vowels used in the present study, the vowel /ae/ is produced with the lowest jaw position and consequently has the largest opening displacement while the vowel /I/ is produced with the highest jaw position and has the smallest displacement. The range of displacements for the three subjects varied considerably, however, the pattern across subjects was the same.

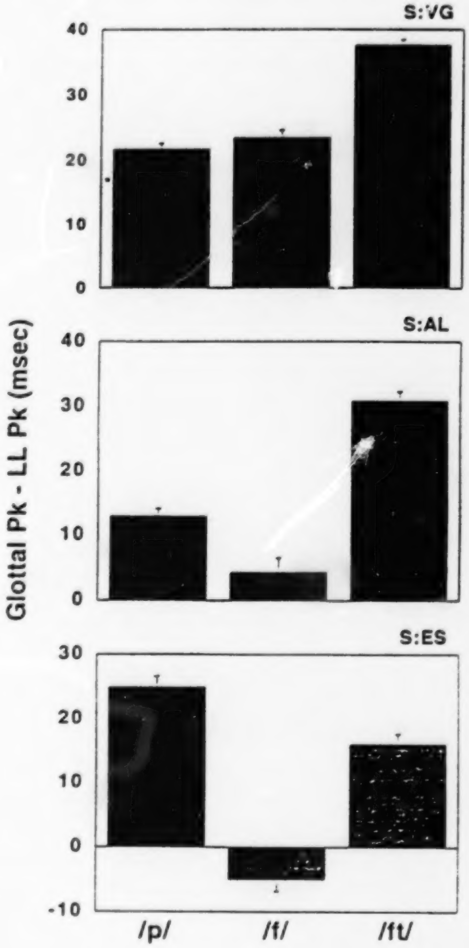


Figure 5. The time interval between the peak glottal opening and lower lip raising for the three subjects. Error bars indicate one standard error. Similar to the results using the time of peak velocity, the positive values indicate that the maximum displacement for lower lip raising occurred before the maximum glottal opening. Only /f/ for S:ES showed a negative value indicating a reversal in the lip-glottal sequence.

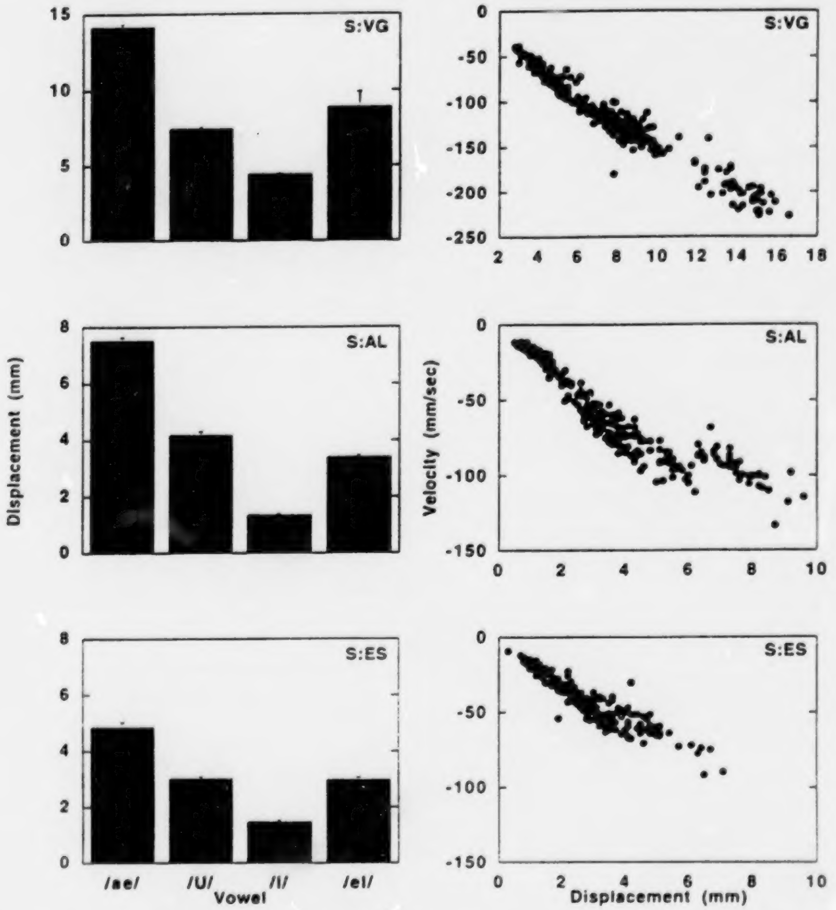


Figure 6. On the left, average jaw lowering displacement (in mm) for the four vowels for the three subjects. Error bars indicate one standard error. On the right, scatterplots of the peak lowering velocity (in mm/sec) as a function of lowering displacement for the four vowels.

It was also found that the lowering motion of the jaw was dependent on the identity of the following consonant. For example, the words "supper" and "suffer" have the same vowel but different following consonants. There was a tendency for the jaw opening displacement for the same vowel to be reduced when followed by /f/ than /p/. This is illustrated in the average lip, jaw and glottal signals presented in Figure 7. Shown are averages (n=40) of the lower lip, jaw, and glottal signals for the utterances "It's a supper" and "It's a suffer" spoken by S:VG. The vertical jaw lowering displacement is reduced and the resulting jaw raising is of greater displacement and higher position when the consonant is /f/ compared to /p/ . As summarized in the top portion of Figure 8, this trend was observed for S:AL but not S:ES. From the middle portion of Figure 8 it can be seen that for the words in which the vowel sound was the same but the consonant was /p/ compared to /ft/ (sipping versus sifting) a similar pattern was observed. In contrast, the bottom portion of the figure illustrates that the jaw opening displacement for the same vowel did not differ when the following consonants was /f/ versus /ft/ (safe versus safety). It should also be noted that for the two subjects that showed a reduction in jaw lowering extent when the following sound was /f/ compared to /p/, a similar reduction was noted for the jaw lowering peak velocity. That is, the reduction in the jaw movement displacement was not due to the raising movement moving closer to the lowering movement and truncating the final displacement. Rather, the jaw lowering motion for a specific vowel was actively adjusted dependent on the identity of the subsequent consonant.

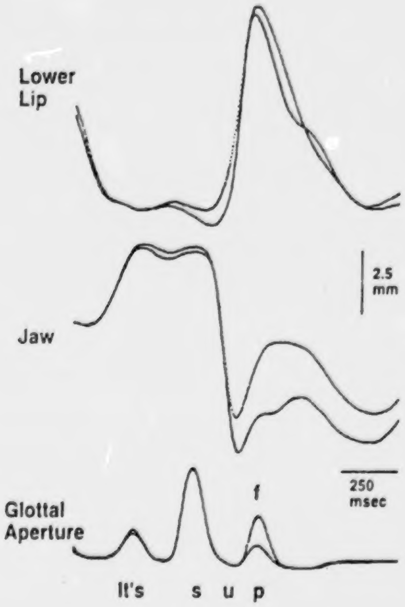


Figure 7. **Averaged lower lip, jaw, and glottal signals (n=40) for S:VG for the phrases "It's a supper" and "It's a suffer"; the dotted line indicates the /f/. Raising motion is up for the jaw and lower lip; increases in glottal aperture is also up. There are two points of interest; the jaw lowering movement is reduced for "u" when the following consonant is /f/ compared to /p/ and the glottal aperture is larger for /f/ than for /p/.**

Oral closing. Further inspection of the average signals in Figure 7 also reflect some additional characteristics of the differences associated with the identity of the oral closing consonant. The extent of lower lip movement for /p/ and /f/ are similar although there appears to be differences in the velocity of the raising. Second, the glottal aperture is larger for /f/ than for /p/. Figure 9 presents the average displacement and peak lower lip raising velocity for the different contexts for the three subjects. The displacement for /p/ and /f/ were generally similar with the exception of the results for S:AL. In contrast, the closing peak velocity was significantly different with /p/ raising velocity higher than /f/ for all subjects. The movement durations were shorter for /p/ than /f/ for all subjects; the average durations were 136, 89, and 89 msec for /p/ and 175, 138 and 140 msec for /f/ for subjects VG, AL, and ES respectively. It can also be seen that the consonant sequence /ft/ produces some changes in the kinematic characteristics of the lip raising action. In general the lip displacement is reduced and the velocity is lower for /ft/ compared to /f/ at least for two of the three subjects.

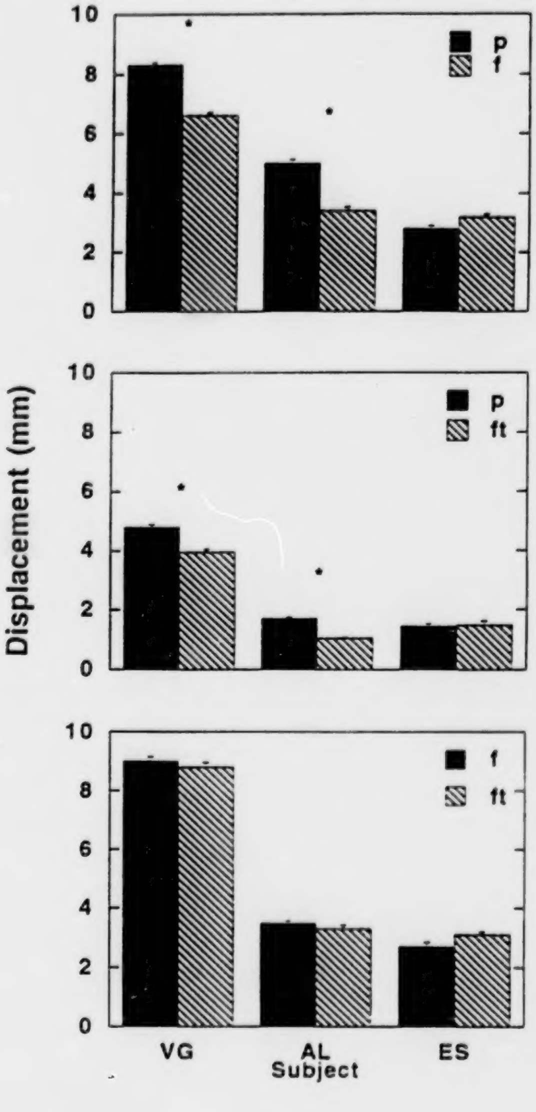


Figure 8. **Average jaw lowering displacement (in mm) for the three subjects comparing the effects of the following consonant on the preceding jaw lowering movement for the same vowel. The top panel contrasts the jaw lowering displacement for the vowel /U/ in the words "supper" and "suffer"; the middle panel contrasts the vowel /I/ before /p/ and /ft/ in the words "sipping" and "sifting"; the bottom panel contrasts the vowel /eI/ before /f/ and /ft/ in the words "safe" and "safety." Error bars indicate one standard error. The differences for S:VG and S:AL for the top and middle comparisons were statistically different ($p < .001$); for S:ES neither comparisons reached significance ($p > .1$). For the /f/, /ft/ comparisons (bottom), there were no significant differences ($p > .1$).**

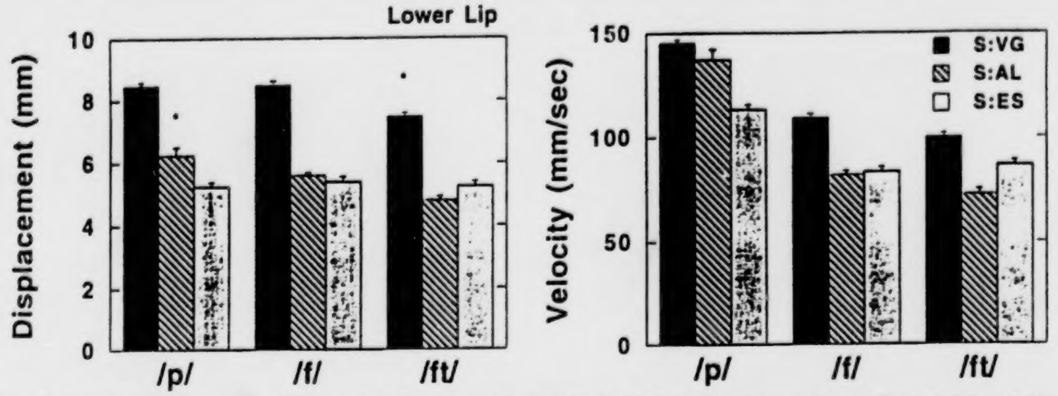


Figure 9. **Lower lip raising displacement (left) and peak velocity (right) for the three subjects for /p/, /f/, and /ft/. There was a slight tendency for a reduction in lower lip displacement for /f/ and/or /ft/ compared to /p/ for two of the subjects (S:VG & S:AL). In contrast, the peak raising velocity demonstrated a robust reduction for /f/ compared to /p/ for all subjects with a smaller difference noted for /ft/ compared to /f/. Error bars indicate one standard error.**

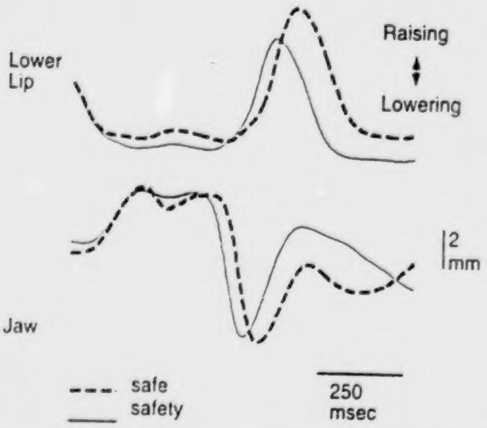


Figure 10. **Average lower lip and jaw vertical movements (n=40) for S:VG illustrating the displacement differences due to the consonant sequence /ft/ compared to /f/. The dotted line indicates the word "safe"; the solid line indicates the word "safety." The lower lip is reduced in raising displacement and increased in jaw displacement for the /ft/.**

Another example of the effect of two consonants in a sequence can be seen in Figure 10. Shown are averages of the lower lip and jaw movements associated with the words "safe" and "safety." The jaw lowering movement for the vowel is similar for the two words. However, the lip and jaw raising movements are different in extent for /f/ compared to /ft/. Since the jaw is involved in elevating the tongue for the /t/ the jaw continues past the position for /f/. As a result the lip raising action is adjusted for the greater jaw contribution to the initial raising. A summary of the lip and jaw contribution to the oral closing is presented in Figure 11. Since it was shown previously that the jaw lowering movement extent for oral opening varied as a function of the vowel identity, it was necessary to normalize the lip and jaw raising to the extent of jaw lowering for each vowel. A gain was derived as the ratio of the jaw lowering displacement to the lip and jaw raising displacement. As can be seen there is a trend for the gain to be higher for /ft/ than for /p/ and /f/ and the gain for /p/ and /f/ are not significantly different. For the jaw, the gain increases from /p/ to /f/ to /ft/ for two of the three subjects (S:VG and S:AL).

All components of the glottal signal were found to differ according to the consonant sequence. The initial analysis focused on the characteristics of the glottal signal for the different consonants. Each glottal action for a voiceless consonant has two distinct phases; an abductory (opening) phase and an adductory (closing) phase. In order to determine whether each phases of the glottal action is an independent action or an interdependent action in which the phases are modulated as a unit, the peak opening and closing glottal velocities for each consonant as well as the initial /s/ were examined. Shown in Figure 12 are scatterplots of the opening and closing velocities for /s/ (left side) and the /p/, /f/ and /ft/ (right side) for the three subjects. The opening and closing velocities for both comparisons systematically covary. From the differences in the data ranges it can be seen that the peak glottal velocity for opening and closing for /s/ was always higher than for any of the other three consonants. To varying degrees it was also the case that the opening and closing velocity exhibited a hysteresis with the opening velocity higher than the closing velocity for /s/ and for the other consonants as a group ($p < .0001$ for all subjects).

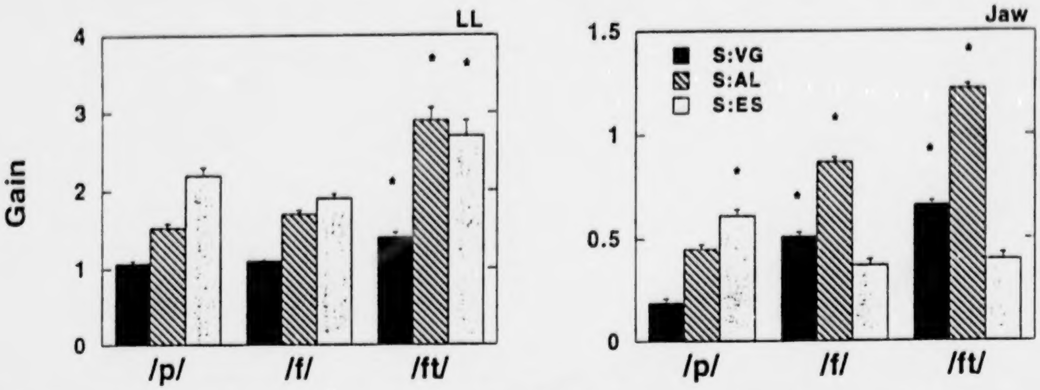


Figure 11. Lower lip and jaw gain defined as the ratio of the raising displacement to the opening displacement for the preceding vowel. Lower lip gain is consistently higher for the consonant sequence /ft/ as is the jaw with the exception of S:ES. In addition, the jaw gain is higher for /f/ compared to /p/ for S:VG and S:AL. Asterisks indicate a significant difference ($p < .01$).

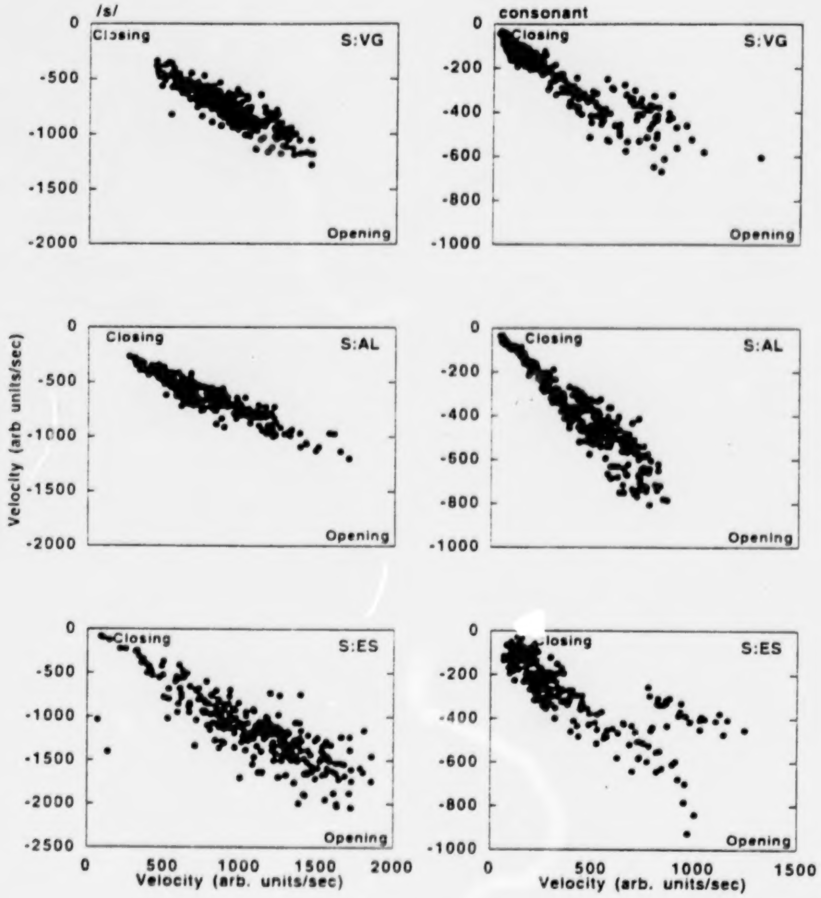


Figure 12. Scatterplots of the glottal opening and glottal closing peak velocity for the /s/ in the different words (left) and the consonants /p/, /f/, and /ft/ (right) for the three subjects. The opening and closing velocities covary strongly for both conditions. Moreover, the values are generally higher for /s/ than for any of the consonants and the opening velocity is generally higher than the closing velocity.

In order to compare the glottal characteristics for the different consonants it was first necessary to amplitude and time normalize the glottal signal. It was reasoned that during the course of the experiment any deviations in the timing or amplitude of the glottal signal unrelated to the phonetic context, such as speaking rate variations or light source changes due to movement of the endoscope, would be evident in the signal for /s/ and normalizing to the /s/ kinematics would minimize any spurious changes. With the exception of the /s/ opening glottal velocity before /ft/ for S:VG there were no significant consonant related differences for either /s/ opening or closing glottal velocity. As such, all the glottal signals for /p/, /f/, and /ft/ were normalized to the glottal signal for the /s/ in each target word. Shown in Figure 13 are the normalized mean glottal opening peak amplitude, duration, and opening and closing velocities for the three subjects. As shown, the glottal aperture is larger, and opens and closes faster for /f/ compared to /p/. Interestingly, the glottal opening movement duration is longer for /f/ consistent with the slower lip raising movement. Apparently the larger glottal opening is a functional adjustment associated with the aerodynamic or kinematic requirements for /f/ that are different than those for /p/. An additional comparison can be made from the figure. The consonant sequence /ft/ (two voiceless consonants) results in a consistent change in the glottal signal. The glottal aperture for the consonant sequence is larger and longer than for /f/ while the opening and closing velocities are lower. It appears that the glottal signal is some form of additive function of two voiceless phonetic segments.

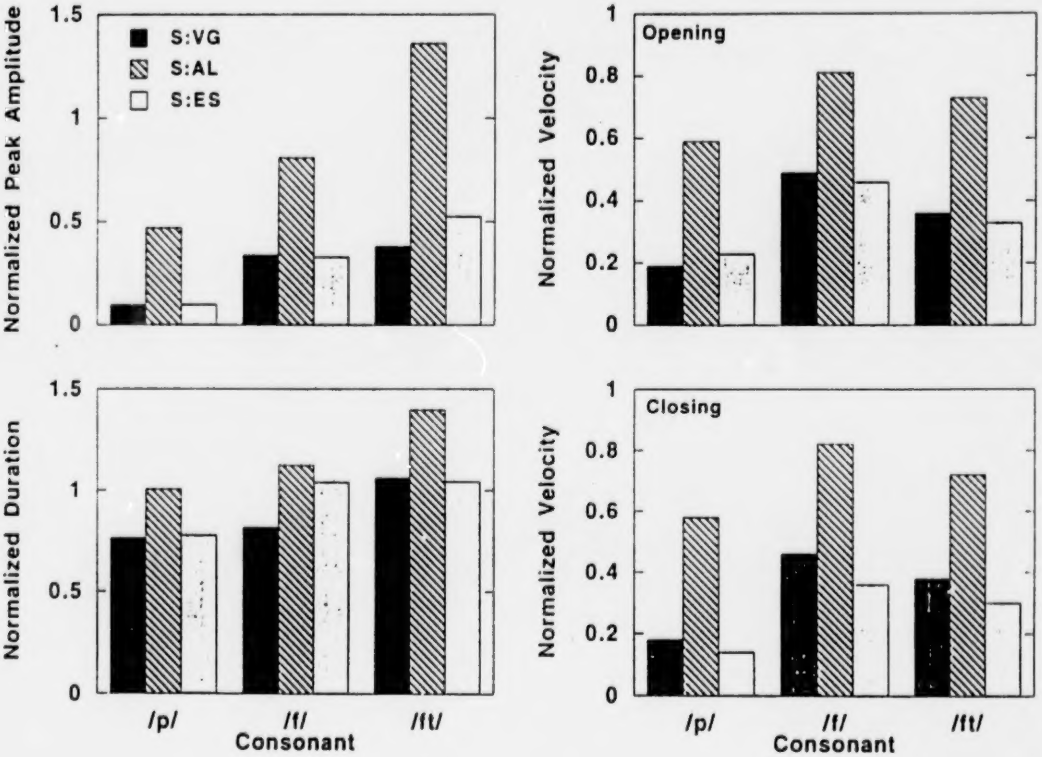


Figure 13. Normalized peak glottal opening, opening duration, and opening and closing peak velocity for the three consonants. The glottal opening is larger and longer in duration for /ft/ compared to /f/ which is larger and longer in duration for /f/ compared to /p/ for all subjects. The opening and closing velocity is greater for /f/ compared to /p/ with a consistent reduction in velocity for /ft/ for the three subjects.

Discussion

Speaking is a sensorimotor process in which cognitive/linguistic intent is transformed into conformational changes in the vocal tract generating the appropriate conditions for the acoustic structure characteristic of a language. This transformation is a time critical process in which multiple muscles and accompanying speech articulators must be coordinated in space and time to produce a variety of vocal tract adjustments. One purpose of the present investigation was to determine whether different articulators cooperating to produce the same sound are coupled in their timing and thereby extend previous observations of speech movement coordination to include an important but relatively inaccessible articulator, the larynx. The results suggest that speech movement timing is a highly systematic and constrained process in which individual articulator actions are controlled as a unit rather than as individual degrees of freedom. A second purpose of the present investigation was to provide detail on the size and characteristics of the underlying units for speech production and to determine the manner in which speech movements are adjusted for phonetic context. Interarticulator timing and the different sound-specific articulator adjustments suggest that speech production units are organized at a level reflecting sound generating segments. Finally, phonetic context was found to produce systematic variations in the relative contribution of the different articulators to the overall movement patterns suggesting an important distinction between the units (speech motor programs) and the adjustments of the units (speech motor programming). These issues will be discussed in the following sections.

Speech movement coordination. The present results extend previous observations on speech movement timing to include the temporal coordination of the lips, jaw, and larynx. These three articulators are critically involved in many of the sounds of English and the significant covariation in their timing reflect some properties of the speech production process. In previous studies it has been shown that the consistency of speech movement timing is partially dependent on the specific articulator action (opening/lowering or closing/raising) which is generally associated with different classes of speech sounds (e.g.,., vowels and consonants; Gracco, 1988; Gracco, 1994). Speech motor actions associated with time critical closing adjustments for certain voiceless consonant sounds, such as /p/ and /f/ in the present study, appear to be highly constrained in their timing. This is apparently to assure that functionally-related actions generate the necessary and sufficient aerodynamic conditions to produce perceptually acceptable acoustic products. In contrast to previous results (Gracco, 1988; Gracco, 1994) which demonstrated a lack of robust relative timing among the lips and the jaw during oral opening, the present results suggest that similar constraints are operating for oral opening actions as well. The difference from previous studies is related to the articulators examined. In the previous studies, the lips and jaw were only examined and the apparent difference between the two general actions appears to be related to different articulators being involved in different linguistic-motor actions that overlap in time (see also Gracco, 1994). At the onset of oral opening the lips are still involved in the consonant sound while the jaw becomes functionally decoupled from the consonant and is directly involved in the following vowel sound. As shown in the present investigation for oral opening, the timing of the jaw and larynx are coupled in their relative timing as their actions are functionally-related to the production of the vowel. The systematic and consistent timing covariation among the articulators for oral opening and closing indicate that timing constraints may be a fundamental property of speech movement coordination. However, it is the case that the relative strength of the coupling varied, with the oral closing actions more highly correlated than oral opening. One possible explanation is that these two actions, oral closing and oral opening, reflect two important but distinct characteristics of speech production. In the case of the oral closing, examination of the relative timing focused on a single speech motor action (the production of a specific phoneme) and the consistent timing of the movements represents the coordination of multiple speech articulators within a specific action unit. In the case of oral opening, examination of the relative timing focused on a transition region between two contiguous speech actions (the transition between a consonant and a vowel). As such, the present investigation examined multiarticulator coordination within a speech production unit and the sequencing of such units into larger aggregates. It is also the case that the oral closing actions were associated with rapid movements and high pressure consonants sounds while the oral opening actions were associated with slower movements and low pressure vowel sounds. The extent to which these factors influence the relative timing among speech articulators is open to empirical investigation. The next section will focus on some of the characteristics of the units for speech production followed by a discussion of the potential mechanisms for sequencing and adjusting the units for phonetic contexts.

Speech motor programs. Speech motor programs can be thought of either as high level goals or procedures for implementation of intent (Schaffer, 1992). A synthesis of these two views can been suggested in which speech motor programs are viewed as neuromuscular configurations that define the structure (intent) of the vocal tract for each unique element (sound) of the language (Gracco, 1990; 1991). Speech motor programs reflect a characteristic neuromuscular configuration that specifies the muscles to be activated and some general characteristics of that activation (Gracco, 1991; Gracco, 1994). Similar to the concept of a motor plan (Evarts et al., 1971) a finite number of such programs would be established during speech motor development and modified periodically for changes in vocal tract shape due to growth. As noted here and elsewhere, speech motor actions appear to be organized at a functional (sound producing) level with control exerted over large regions of the vocal tract rather than over the action of individual articulators (Gracco & Abbs, 1986; Gracco, 1990; 1991). The present results are consistent with this conception in that for the three different consonant sounds examined (/s/, /p/, /f/) the articulatory configurations were unique and significantly different along a number of kinematic dimensions. Based on these and previous results demonstrating the consistent relative timing among functionally-related articulator coordination it is further suggested that an important component of each speech motor program is the relative timing among the neuromuscular elements. Rather than explicitly controlling the timing among articulators such as the lips, jaw, and larynx, their coordination is an inherent component of the unit (program). These learned motor programs are stored in memory and provide the physiological framework for the sounds of the language reflecting the physiological instantiation of the phoneme.

Speech motor programming. One of the criticisms with the construct of motor programs underlying voluntary behavior is the lack of adaptability often cited as a limitation for such a metaphor (Kugler & Turvey, 1987; Kelso, 1986). The lack of adaptability is of some significance since it is well known that the phonetic context of a particular sound can substantially modify its peripheral (kinematic and acoustic) manifestations. The same sound produced at the beginning versus the end of a syllable and between different vowels will display different movement patterns. The widespread presence of contextual variation has even led to an extreme, though currently unpopular view, that all possible variations of the sounds of the language are stored in memory as part of the speech coding process (Wickelgren, 1969). Based on the results from the present investigation, and results from investigations of the sensorimotor mechanisms of speech motor control, a more realistic perspective can be presented. A number of investigations have demonstrated that mechanical perturbations to the lips and jaw result in short latency (within a reaction time) responses in all the articulators activated for the specific speech sounds (Abbs & Gracco, 1984; Folkins & Abbs, 1975: Gracco & Abbs, 1985; Kelso et al., 1984: Shaiman, 1989). It appears that somatic sensory receptors located within the vocal tract have the requisite properties to interact with the central motor commands to provide adaptive adjustments in the speech motor programs resulting from peripheral variations in phonetic context (Gracco, 1987; Gracco & Abbs, 1988). On line sensorimotor mechanisms provide one means to adjust central commands to changes in peripheral conditions. The present results also suggest that an additional central mechanism is operating for contextual adjustments. For two of the three subjects it was shown that the jaw lowering extent for the vowel /U/ was affected by the identity of the following consonant. When the following consonant was /f/ the jaw lowering movement was reduced in amplitude compared to when the following consonant was /p/. This affect was not merely the result of the consonantal raising movement truncating the jaw lowering movement since the jaw lowering velocity was also reduced in the /f/ context. This phenomenon of coarticulation, or the anticipatory modification of speech output due to context, is of some neurophysiological significance. It suggests that what is to be said is planned in advance and the overall context can influence aspects of the central commands. These two complementary processes operating on a framework of learned motor programs provide the flexibility characteristic of speech production. It suggests that a distinction can be made between speech motor programs, as goal directed phonetically-based actions, and dynamic (programming) processes that provide adaptive and on-line adjustments to speech motor sequences. Moreover, speech motor adjustments associated with the consonant were distributed to the preceding vowel action suggesting that the speech motor programming operates over an interval on the order of at least a movement cycle involving two or more phonetic segments (see also Gracco, 1994).

Speech motor sequences. The present investigation allowed an examination of the kinematic effects of two consonants produced in sequence. The consonant sequence /ft/ involves two voiceless sounds that overlap. Functionally the consonant sequence /ft/ requires the lip to contact the upper teeth followed by tongue tip contact with the roof of the mouth for the /t/. As shown in Figure 10 comparing /safe/ with /safety/ the jaw position is higher for /safety/. It can also be seen that the jaw moves continuously from a minimum for the vowel to some maximum value associated with the /t/. The jaw position for the /f/ in the /ft/ sequence was not unambiguously identifiable. This is a characteristic of many speech motor actions and is the basis for the difficulty in segmenting continuous motion into the underlying discrete units. While the jaw passes through some spatial target for /f/ it does not (and need not) stop its motion. It is suggested from jaw movement considerations that two successive target positions are reflected in the single trajectory (see also Flanagan, Ostry, & Feldman, 1993 for arm movements to displaced targets). For the larynx, the consonant sequence also resulted in an apparent blending of the two voiceless consonants with the resulting glottal amplitude and/or duration for /ft/ larger and/or longer than /f/. As such, the consonant sequence produced a hybrid pattern adjusted to accommodate the longer duration voiceless segment.

Conclusions

The present investigation was initiated to evaluate the coordination and motor control for speech by examining the interactions of the lips, jaw, and larynx in different phonetic contexts. While limited in scope the present results suggest a number of general properties of speech production and its motor control. The timing among functionally-related articulators suggest that speech movements are organized into aggregates larger than individual articulators. The different kinematic patterns associated with the different consonant and vowel sounds examined in the present investigation further suggests that different sounds have different neuromotor specifications. It appears that each sound in the language has associated with it a neuromotor representation reflecting the muscles to be activated and their specific spatiotemporal coordination. These fundamental units (speech motor programs; coordinative structures) provide the framework for speech production and appear to reflect the neurobiological equivalent of the phoneme. Additional modulatory processes exist to scale and sequence the phonetic units into larger sequences for communication (syllables, words, phrases, etc.) by adjusting the vocal tract characteristics over an interval larger than individual phonemes (phonetic segments). As suggested recently, the unit of programming is, minimally, on the order of a movement cycle (or syllable) and within this interval contextual variations are adjusted based on the immediate state of the vocal tract and the compatibility of the neighboring sounds (Gracco, 1994). The dynamic nature of speech production results in blending of movements that modify the peripheral manifestation of the underlying units and obscures their identification. As such, the neural control specifications of the units must be sufficiently relaxed to allow for contextual variations. This is also reflected in the ability of the listener's perceptual system to handle the lack of invariance and maintain highly reliable information transfer.

REFERENCES

Abbs, J. H., & Gracco, V. L. (1984). Control of complex motor gestures: Orofacial muscle responses to load perturbations of the lip during speech. *Journal of Neurophysiology, 51(4)*, 705-723.

Baer, T., Löfqvist, A., & McGarr, N. S. (1983). Laryngeal vibrations: a comparison between high-speed filming and glottographic techniques. *Journal of the Acoustical Society of America, 73*. 1304-1308.

Bernstein, N. (1967). *The co-ordination and regulation of movements*. New York: Pergamon Press.

Evarts, E. V., Bizzi, E., Burke, R., & DeLong, M. (1971). Central control of movement. *Neuroscience Research Program Bulletin, 6*, 1-70.

Flanagan, J. R., Ostry, D. J, & Feldman, A. G. (1993). Control of trajectory modifications in target-directed reaching. *Journal of Motor Behavior, 25*, 140-152.

Folkins, J. W, & Abbs, J. H. (1975). Lip and jaw motor control during speech: responses to resistive loading of the jaw. *Journal of Speech and Hearing Research, 18*, 207-220.

Gracco, V. L. (1994). Some organizational characteristics of speech movement control. *Journal of Speech and Hearing Research, 37*, 4-27.

Gracco, V. L. (1991). Sensorimotor mechanisms in speech motor control. In, H. Peters, W. Hultsijn, & C. W. Starkweather (Eds.), *Speech motor control and stuttering* (pp. 53-78). North Holland: Elsevier.

Gracco, V. L. (1990). Characteristics of speech as a motor control system. In G. Hammond (Ed.), *Cerebral control of speech and limb movements* (pp. 3-28). North Holland: Elsevier.

Gracco, V. L. (1988).Timing factors in the coordination of speech movements. *Journal of Neuroscience, 8*, 4628-4634.

Gracco, V. L. (1987). A multilevel control model for speech motor activity. I H. Peters & W. Hulstijn (Eds.), *Speech motor dynamics in stuttering* (pp. 57-76). Wien: Springer-Verlag.

Gracco, V. L., & Abbs, J. H. (1988). Central patterning of speech movements. *Experimental Brain Research, 71*, 515-526.

Gracco, V. L., & Abbs, J. H. (1987). Programming and execution processes of speech motor control. Potential neural correlates. In E. Keller, & M. Gopnick (Eds.), *Motor and sensory processes of language* (pp. 163-202). Hillsdale, NJ: Lawrence Erlbaum.

Gracco, V. L., & Abbs, J. H. (1986). Variant and invariant characteristics of speech movements. *Experimental Brain Research, 65*, 156-166.

Gracco, V. L., & Abbs, J. H. (1985). Dynamic control of the perioral system during speech: kinematic analyses of autogenic and nonautogenic sensorimotor processes. *Journal of Neurophysiology, 54*, 418-432.

Kelso, J. A. S. (1986). Pattern formation in speech and limb movements involving many degrees of freedom. In H. Heuer & C. Fromm (Eds.), *Generation and modulation of action patterns* (pp. 105-128). Berlin: Springer-Verlag.

Kelso, J. A. S., Tuller, B., V.-Bateson, E., & Fowler, C. (1984). Functionally specific articulatory cooperation following jaw perturbations during speech: Evidence for coordinative structures. *Journal of Experimental Psychology: Human Perception and Performance, 10*, 812-832.

Kollia, H. B., Gracco, V. L., & Harris, K. S. (submitted). Lip, jaw, velar coordination during speech. *Journal of the Acoustical Society of America.*

Kugler, P. N. & Turvey, M. T. (1987). *Information, natural law and the self-assembly of rhythmic movement.* Hillsdale: Lawrence Erlbaum.

Lacquaniti, F. & Soechting, J. F. (1982). Coordination of arm and wrist motion during a reaching task. *Journal of Neuroscience, 2*, 399-408.

Löfqvist, A., & Yoshioka, H. (1980). Laryngeal activity in Swedish obstruent clusters. *Journal of the Acoustical Society of America, 68*, 792-801.

Macchi, M. (1988). Labial articulation patterns associated with segmental features and syllable structure in English. *Phonetica, 45*, 109-121.

MacKenzie, C. L. (1992). Constraints, phases and sensorimotor processing in prehension. In G. E. Stelmach & J. Requin (Eds.), *Tutorials in motor behavior.* (pp. 181-194). Amsterdam: Elsevier Science Publishers B. V.

MacPherson, J. M. (1988a). Strategies that simplify the control of quadrupedal stance. I. Forces at the ground. *Journal of Neurophysiology, 60*, 204-217.

MacPherson, J. M. (1988b). Strategies that simplify the control of quadrupedal stance. II. Electromyographic activity. *Journal of Neurophysiology, 60*, 218-231.

Oshima, K., & Gracco, V. L. (1992). Mandibular contributions to speech production. *Proceedings of the Conference on Spoken Language Processing*, 775-778.

Schaffer, L. H. (1992). Motor programming and control. In G. E. Stelmach & J. Requin (Eds.), *Tutorials in motor behavior.* (pp. 181-194). Amsterdam: Elsevier Science Publishers B. V.

Shaiman, S. (1989). Kinematic and electromyographic responses to perturbation of the jaw. *Journal of the Acoustical Society of America, 86*, 78-87.

Soechting, J. F., & Lacquaniti, F. (1989). An assessment of the existence of muscle synergies during load perturbations and intentional movements of the human arm. *Experimental Brain Research, 74*, 535-548.

Turvey, M. T. 1977. Preliminaries to a theory of action with reference to vision. In, R. Shaw & J. Bransford (Eds.), *Perceiving, acting and knowing: Towards an ecological psychology.* Hillsdale: Lawrence Erlbaum.

Wickelgren, (1969). Context-sensitive coding, associative memory, and serial order in (speech) behavior. *Psychological Review, 76*, 1-15.

FOOTNOTE

*To appear in *Journal of Neuroscience* (1994).

Haskins Laboratories Status Report on Speech Research
1993, SR-115/116, 33-44

An Unsupervised Method for Learning to Track Tongue Position from an Acoustic Signal*

John Hogden,† Philip Rubin,‡ and Elliot Saltzman

A procedure is demonstrated for learning to recover the relative positions of simulated articulators from speech signals generated by articulatory synthesis. The algorithm learns without supervision, that is, it does not require information about which articulator configurations created the acoustic information in the training set. The procedure consists of vector quantizing short time windows of a speech signal, then using multidimensional scaling to represent quantization codes that were temporally close in the encoded speech signal by nearby points in a *continuity map*. Since temporally close sounds must have been produced by similar articulator configurations, sounds which were produced by similar articulator positions should be represented close to each other in the continuity map. Continuity maps were made from parameters (the first three formant center frequencies) derived from acoustic signals produced by an articulatory synthesizer that could vary the height and degree of fronting of the tongue body. The procedure was evaluated by comparing estimated articulator positions with those used during synthesis. High rank-order correlations (0.95 to 0.99) were found between the estimated and actual articulator positions. Reasonable estimates of relative articulator positions were made using 32 categories of sound and the accuracy improved when more sound categories were used.

1. INTRODUCTION

A growing body of research (e.g., Atal, 1975; Boe, Perrier, & Bailly, 1992; Hogden, Löfqvist, Gracco, Oshima, Rubin, & Saltzman, 1993; Jordan & Rumelhart, 1992; Kawato, 1989; Kuc, Tutuer, & Vaisnys, 1985; Ladefoged, Harshman, Goldstein, & Rice, 1978; McGowan, 1994; Papcun, Hotchberg, Thomas, Laroche, Zacks, & Levy, 1992; Rahim, Kleijn, Schroeter, & Goodyear, 1991; Schroeter & Sondhi, 1992; Shirai & Kobayashi, 1986) supports the hypothesis that information about articulator positions can be recovered from the acoustic speech signal. This conclusion is somewhat surprising since, when the acoustic properties of the vocal tract are modeled by lossless acoustic tubes, radically different vocal tract shapes can have identical transfer functions (Fant, 1970; Flanagan, 1972). Furthermore, although adding a glottal energy loss to the vocal tract model can make the mappings from acoustics to vocal tract shapes unique (Markel & Gray, 1976), adding energy losses is not always sufficient to eliminate vocal tract shape ambiguities (Atal, Chang, Matthews, & Tukey, 1978)

Energy losses or not, it is clear that the shape of an acoustic tube cannot be uniquely determined from information about formant frequencies of a single transfer function without incorporating additional constraints. This has been shown using articulatory synthesizers, both with and without energy losses (Atal, et al., 1978; Maeda, 1989; Stevens & House, 1955). Linear prediction theory leads to the same conclusion by showing that formant frequencies and bandwidths must both be used to determine vocal tract shape. Finally, bite-block experiments confirm that people can produce vowels with nearly normal values of the first three formant frequencies using a "physiologically unnatural position of the mandible" (Lindblom, Lubker, & Gay, 1979). It is difficult to argue that bite block vowels are acoustically identical to normally produced vowels—perceptual differences between normal and bite-block vowels have been noted (Fowler &

This research was supported by NIH Grant DC-00016 and NIH Grant DC-00121 to Haskins Laboratories.

Turvey, 1980)—but Lindblom et al. found that the first three formants of bite block vowels were usually within 3 standard deviations of normal vowels formants with few systematic deviations.

Nonetheless, there has been some success at recovering articulation from acoustics. For example, given a training set consisting of acoustic signals generated by an articulatory synthesizer and the articulator positions used to produce them, Atal (1975) found a non-linear regression function that calculated seven vocal tract parameters (constriction location, constriction degree, lip protrusion, etc.) from twelve acoustic parameters (six formant frequencies and six bandwidths). The importance of using as much acoustic information as possible was reinforced in this study because Atal was able to determine vocal tract parameters from representations of the acoustic signal provided the acoustic representation included information about a sufficient number of formant frequencies and bandwidths.

Atal's success in the 1975 study was based, at least in part, on the fact that the model vocal tract used for synthesis had fewer degrees of freedom than the acoustic information used to recover the tract shape. Conversely, the finding by Atal et al. (1978) that many different vocal tract shapes can lead to the same acoustic signal was partly due to the fact that the number of articulatory parameters to recover was greater than the number of acoustic parameters measured. Clearly, vocal tract shape can be determined more accurately if the number of acoustic parameters used to determine vocal tract shape exceeds the number of parameters used to describe vocal tract shape.

Unfortunately, as Sondhi (1979) mentions, only a limited number of acoustic parameters can be accurately recovered from speech. This poses the serious question of whether the vocal tract shapes used during speech can be described with fewer parameters than the number of acoustic parameters that can be accurately recovered from speech. There is support for the contention that the articulator positions commonly used during vowel production can be described parsimoniously; tongue shape can be adequately represented by only 2 or 3 parameters (Harshman, Ladefoged, & Goldstein, 1977; Morrish, Stone, Shawker, & Sonies, 1985) and vocal tract shapes in general can be represented by about 7 to 10 factors (Coker, 1976; Maeda, 1989; Rubin et al., 1981). Evidence that human articulator positions can be recovered from acoustic information has been presented by Ladefoged et al. (1978), who used multiple regression to find a relationship between the first three formants and two PARAFAC factors representing tongue shape. Tongue positions inferred from the first three formants of steady state vowels accurately reflected the tongue positions seen in X-ray tracings for several subjects, although there was some difficulty in estimating the tongue shapes used to produce the vowel [a]. Similarly, Hogden et al. (1993) recovered articulator positions using a look-up table.

Some articulatory features can be more easily recovered from speech than others. For example, Boe et al. (1992) used an articulatory synthesizer based on X-ray data (Maeda, 1979) to show that the location and area of the oral constriction used in vowel production could be determined from the first three formants alone, even though the complete shape of the vocal tract could not be recovered. This research demonstrates that even if the vocal tract shape is not entirely recoverable from the acoustic signal, aspects of articulation that are important for phonetic identification may be recoverable. Continued research in this direction may uncover other articulatory features that can be determined despite ambiguous mappings from acoustics to vocal tract shape.

Most techniques for solving the acoustic-to-articulatory mapping problem have not been rigorously tested on human articulatory/acoustic data because of the difficulties involved in measuring the articulator positions. Three exceptions to this rule are the studies by Ladefoged et al. (1978) and Hogden et. al. (1993) that were already discussed, and also a study by Papcun et al. (1992). The latter study found that neural networks, which perform a type of non-linear regression, can calculate X-ray microbeam pellet positions from spectral information. As in Atal's nonlinear regression study, Papcun et al. supplied their recognition algorithm with more acoustic information than simple measurements of formant frequency. One difference between Atal's study and the study by Papcun et al. is that Atal used acoustic signals from static vocal tracts while Papcun et al. gave the neural network spectral information from successive short-time windows of speech, essentially providing acoustic information from successive vocal tract shapes. This difference is important because using information from several spectral slices can help overcome one-to-many mapping problems (Kuc, et al., 1985; Rahim, et al., 1991).

We will describe a novel method, the *continuity mapping* technique (Hogden, 1991; Hogden et al., 1992a), for computing articulator information from the speech wave. The goal of the continuity mapping algorithm is to produce a map, called a continuity map (CM), in which acoustic signals that are produced close together in time are represented by points that are close to each other in the continuity map. The reasoning behind this is that speech sounds mapped to nearby locations in the continuity map (those produced close together in time) must have been produced by similar articulator configurations. We know that temporally proximate acoustic signals were produced by similar articulator configurations because the articulators move continuously, i.e. they do not move from one position to another without occupying intermediate positions. Since sounds produced by similar articulator configurations are mapped close together in the continuity map, the continuity map should give topologically accurate information about articulator positions.

CMs differ from other topological maps of acoustic signals (Kohonen, 1988) in that for CMs acoustic signals are not placed close together on the basis of acoustic similarity. Unlike other topological mapping procedures, the CM algorithm is trying to recover information about articulator positions, and acoustic signals which are completely different can be produced from very similar articulator configurations. For example, the tongue only needs to move a small distance to change from producing a non-fricative to a fricative—drastically different sounds. To recover articulatory information, acoustically dissimilar sounds need to be able to be placed close to each other in the map. By placing acoustics signals close to each other in the CM if they were produced close to each other in time, drastically different acoustics signals can be represented next to each other, something that is not possible when using an acoustic distance measure.

Unlike previous techniques for recovering articulator positions, which determine the absolute positions of the articulators, continuity mapping only determines their relative positions. However, the relative articulator positions are estimated by an unsupervised algorithm, i.e. without giving the algorithm access to explicit information about the articulator positions used to generate the acoustic signals in the training set. Understanding the difference between a supervised and an unsupervised learning algorithm is essential for evaluating the advantages and disadvantages of the continuity mapping algorithm, thus we will discuss it in a little more detail.

Regression can be thought of as a supervised learning technique for estimating y values from x values. To estimate values of y from values of x we find the regression line relating y and x. To calculate the regression line, examples of (x,y) pairs are needed. The best fitting line cannot be found from x values alone. That is the defining characteristic of a supervised algorithm: examples of both the inputs and outputs are needed for learning. Being supervised algorithms, previous methods for determining articulator positions from acoustics require simultaneous measurements of articulator positions and the resulting acoustics.

The continuity mapping algorithm is an unsupervised algorithm. To continue the analogy to regression, using the continuity mapping technique is somewhat like finding the regression line relating x and y when given only the x values. An unsupervised algorithm is not given the desired output values—even during training. If the continuity mapping procedure is successful, it could learn to relate acoustics to articulation from a tape recording of an individual's speech—without any articulatory measurements. In the present work note that, although we do have simultaneous measurements of acoustics and articulation, the continuity maps are made from the acoustic data alone. The articulatory data is only needed to compare estimated articulator positions to the actual articulator positions.

From the above discussion, it should be clear that supervised learning algorithms will be difficult to apply to the problem of recovering articulator positions from acoustics. The difficulty lies in the fact that, to use a supervised learning algorithm to recover articulator positions, we need to gather a huge set of simultaneous articulatory and acoustics data. Without the simultaneous data, the supervised algorithms can not learn to relate acoustics to articulation. Needless to say, it is still very difficult to gather such data, so supervised algorithms are not yet practical solutions to real world problems.

Supervised algorithms are also problematic if you believe that perceiving speech is tantamount to perceiving articulator gestures (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Liberman & Mattingly, 1985). After all, when children perceive speech produced by others, the children are not told what articulator positions the other speakers are using. While the children do have access to information about their own articulations, a child's speech is acoustically different from adult speech, so it is difficult to imagine how the child could *learn* to relate adult acoustics to articulator positions given only examples of child speech (although innate knowledge, or possibly some kind of normalization, could be used to get around this problem). Similarly, it is difficult to understand how people could learn to perceive sounds which they cannot produce, as in sounds from foreign languages, or sounds that a child has not yet learned to produce (Smith, 1973).

Being an unsupervised algorithm, continuity mapping avoids the previously mentioned problems inherent in supervised algorithms; however, some information is lost to gain the advantages of unsupervised learning. Unlike supervised algorithms, the continuity mapping algorithm is not able to recover the absolute positions of the articulators—only the *relative* positions of the articulators can be estimated. Any rotation, reflection, translation, scaling or other topological transformation of the estimated positions will be an equally acceptable solution as far as the continuity mapping algorithm is concerned.

The continuity mapping algorithm also faces normalization problems, i.e. a map relating acoustics to articulation created for one speaker may not be accurate for a different speaker. So, for the continuity mapping algorithm to be useful, we will either need to determine some way to normalize speech signals from different speakers (as is also the case for supervised algorithms), or we will need to make a variety of continuity maps to accommodate different speakers.

Because of the potential advantages of continuity mapping, several continuity maps were created and tested on acoustic data generated by an articulatory synthesizer. The following discussion describes these experiments.

2. GENERATING AN ARTICULATOR MAP

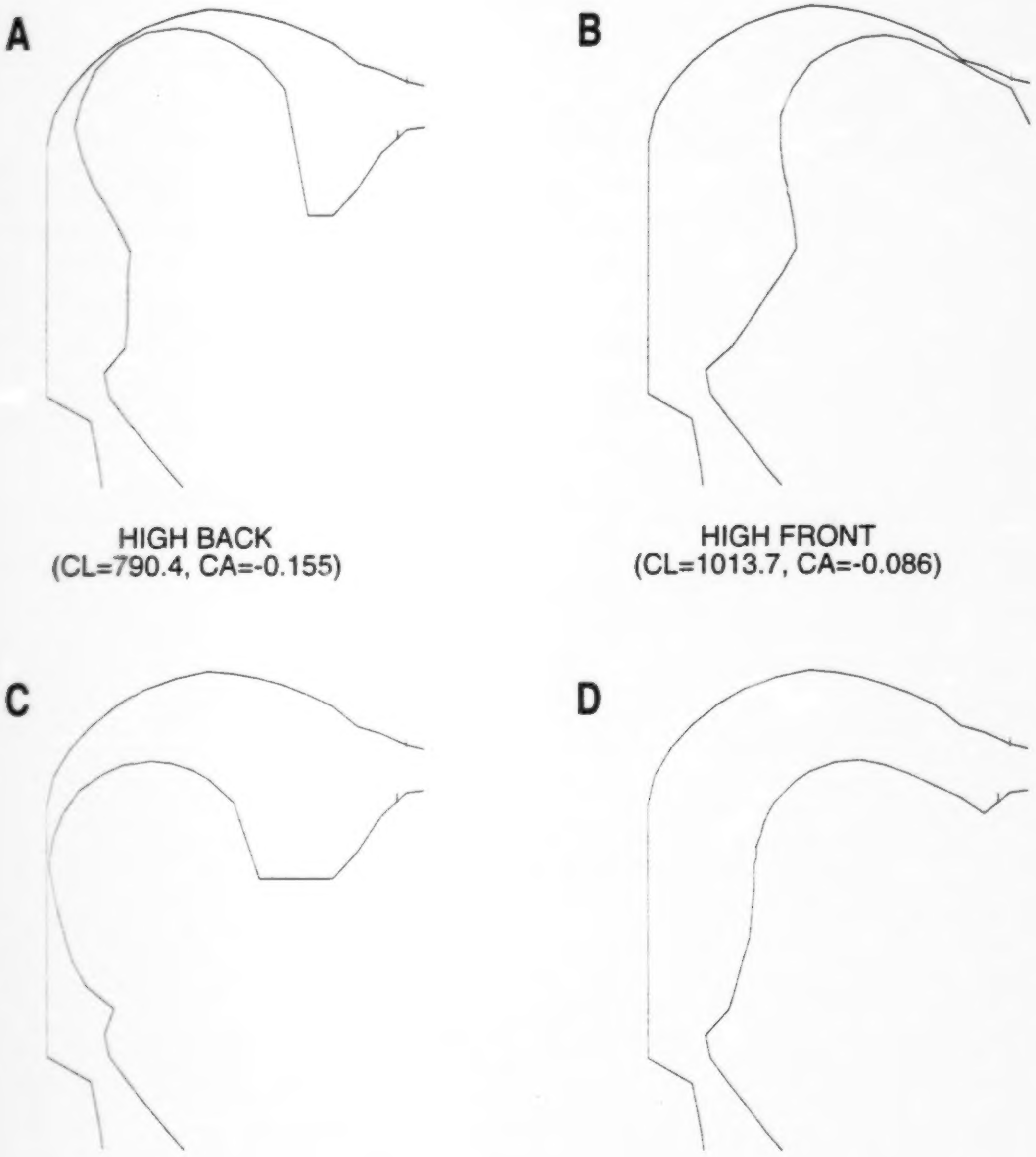
Since gathering simultaneous information about the entire set of articulator positions (especially the tongue) and speech acoustics is quite difficult, the articulatory speech synthesizer at Haskins laboratories (Mermelstein, 1973; Rubin, et al., 1981) was used to generate acoustic signals from static vocal tract configurations. Only the two-dimensional articulator space defined by the synthesizer's degrees of freedom for tongue body motion was investigated. The rest of the articulators were fixed at their neutral positions.

We chose to use two degrees of freedom purely for purposes of illustration. As long as the mapping from acoustics to articulation is not one-to-many, it should be possible to recover information about more than two degrees of freedom as well. However, with articulatory synthesizers, there are typically one-to-many mappings from acoustics to articulation in static synthesis. To stick to our main objective—illustrating the continuity mapping algorithm—our initial work has been limited to recovering two degrees of freedom.

To cover the full range of tongue body positions, the tongue body center was placed at each of 2500 equidistant points in a square grid. Excluding tongue positions that completely closed the vocal tract left 2011 viable tongue positions. Figure 1 gives a flavor of the range of tongue positions by showing some of the more extreme positions. A vector composed of the first three formants of the resulting acoustic signal was calculated for each tongue position.

Each formant vector was replaced by a scalar code using a nearest neighbor coding technique. In nearest neighbor coding, the acoustic similarity between each formant vector and each of a set of prototypical formant vectors is calculated (by finding the Euclidean distance between formant vectors, for example), and the formant vector is replaced by the code representing the most similar prototype. We used a weighted Euclidean distance in formant space as the measure of acoustic similarity. The weight on any formant was the inverse of the standard deviation of the formant, calculated over all tongue positions. The weighted Euclidean distance measure is only one of a variety of distance measures that would all be reasonable. The appropriate distance measure to use for natural speech will likely be more complex (Schroeter, Meyer, Parthasarathy, 1990), but our goal is to illustrate the continuity mapping procedure, so a more complete discussion of possible distance measures is beyond the scope of this paper.

The set of prototypical acoustic signals used in the nearest neighbor coding scheme were derived using a K-means vector quantization (VQ) algorithm (Gray, 1984; O'Shaughnessy, 1987). The VQ algorithm starts with some initial set of acoustic prototypes and moves them around in formant space to minimize the sum of the acoustic distances between the sounds being categorized and the prototypes they are closest to. Since the VQ minimization technique can run into local minima, it needs to be used with different sets of initial prototype positions. We generated three sets of 32 prototypes (each set is called a codebook because the prototypes are referred to by a number called a code) and used for further study the codebook which best minimized the error function.



TONGUE POSITION EXTREMA

HIGH BACK
(CL=790.4, CA=-0.155)

HIGH FRONT
(CL=1013.7, CA=-0.086)

LOW BACK
(CL=826.3, CA=-0.372)

LOW FRONT
(CL=1027.8, CA=-0.228)

Figure 1. **Examples of vocal tract shapes created using extreme tongue positions. The articulatory synthesizer parameters used to make these tongue positions are given in parentheses below the shapes.**

The effect of quantizing the acoustic parameters can be seen in what we call an *articulator map* (AM), like that in Figure 2. Figure 2 shows which vector quantization prototype was most similar to the formant vector produced with each tongue position. The axes of the plot represent tongue body height and frontness and the numbers plotted in the figure are the codes representing the closest prototype. Each code is plotted near the center of a small region which we call an *isocode region*. As the name implies, all the acoustic signals produced with the tongue body positioned within a single isocode region are represented by the same code. It is important to realize that the VQ algorithm generates categories based on acoustics alone. We are able to make the map shown in Figure 2 because, in this experiment, we know the both articulator positions and the resulting derived acoustic values. However, the articulator positions are not needed to perform VQ, and being able to draw the articulator map is not essential for creating a continuity map—the articulator map merely helps to visualize how the continuity mapping algorithm works.

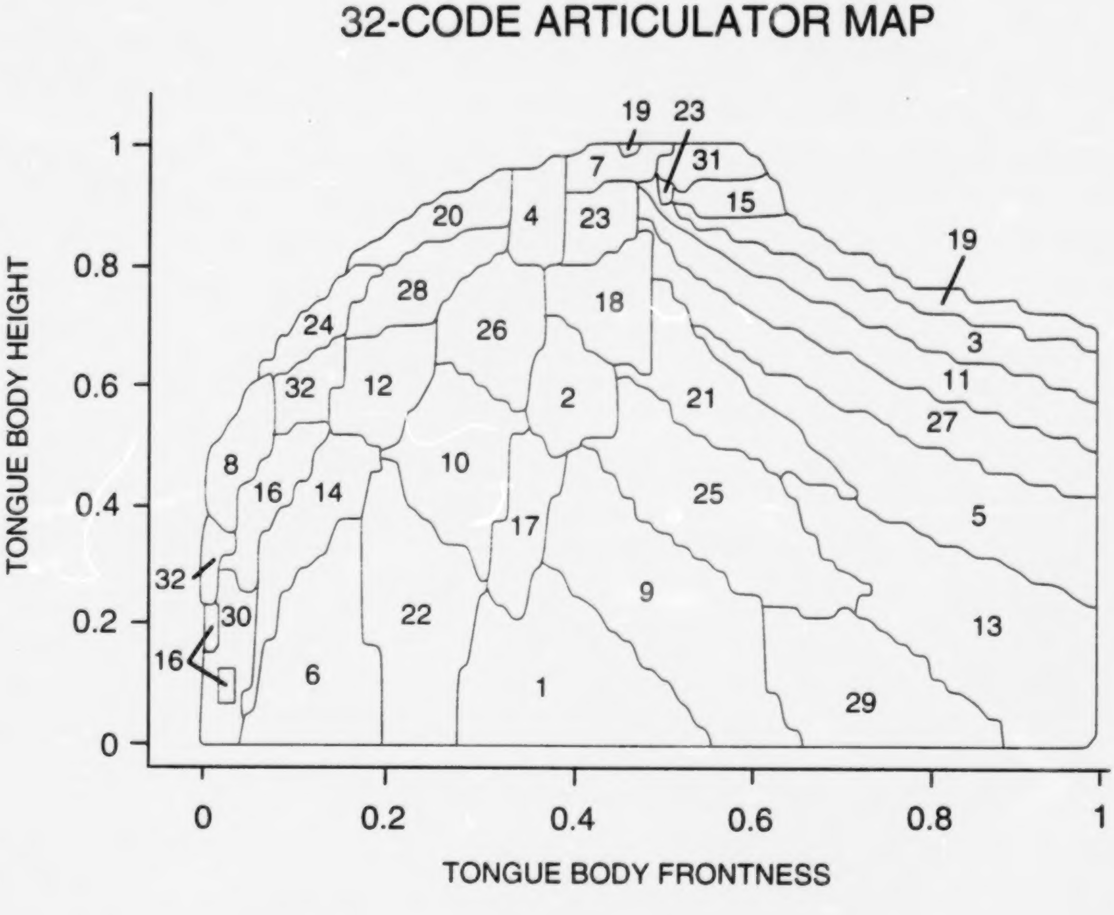


Figure 2. **Articulator map constructed using 32 codes. Each position in the map represents a tongue position. The numbers plotted are codes indicating which acoustic parameters are produced within the isocode regions.**

Notice that some of the codes are produced in two or more distinct regions. Code 16 is one example. As in other studies (Stevens & House, 1955), the codes that occur in disjoint regions are found mostly when the tongue body center is low and back. When a code is found in more than one distinct region, the regions are fairly close together, so the first three formant frequencies are sufficient to determine two tongue position parameters with a relatively small error. As has already been discussed, for synthesized speech the extent of the one-to-many mapping problem depends on the relative number of articulatory and acoustic parameters. Thus, if the first three formants were used to recover more than two articulator parameters, there would probably be more cases where different articulator positions created similar acoustic signals. While we do not want to draw conclusions about the human acoustic to articulatory mapping from this example of synthesized speech, it will be seen that one-to-many mappings do not always prevent good articulator position estimates.

3. GENERATING A CONTINUITY MAP

To allow the continuity mapping algorithm to use information about which signals can be produced close together in time, sequences of codes were produced by taking random walks among the 2011 viable articulator positions. These walks were intended to provide examples of the sounds which could be produced by varying the tongue position in a continuous fashion, so the steps were made short enough to ensure that transitions only occurred between adjacent regions. At each time step, the code produced using the current tongue position was output and the tongue was moved a short distance in some random direction. The random walk continued until at least N transitions were made to each code, with N taking on the value of 1, 2, 3, 4, 5, 10, or 50. Three random walks were made for each value of N, for a total of 21 random walks. As discussed below, random walks provide relatively poor information about the distances between isocode regions, and so give us a conservative means for determining how well the continuity maps will be able to estimate articulator positions.

A set of intercode distance estimates was made for each random walk by calculating the average number of transitions between codes. The average number of transitions is calculated after first eliminating adjacent repetitions of codes, e.g., a sequence like "25, 25, 25, 13, 13, 13, 5, 21, 21, 25, 29, 13, 29, 9" is reduced to "25, 13, 5, 21, 25, 29, 13, 29, 9". The next step is to count the number of transitions between each pair of codes in the sequence. To do this, we start with the first code in the sequence and count the number of transitions to codes occurring later in the sequence. Then we start from the second code in the sequence and count the transitions from there, etc.

Counting from a particular starting code continues until *any* of the codes in this counting sequence is encountered twice. The justification for restarting at code repetitions is that, without restarting, intercode distance estimates would be overestimates. In the example given, the distance between code 25 and code 29 is overestimated by counting the number of transitions from the initial 25 to the 29 because the second occurrence of code 25 is adjacent to code 29. So, starting from the initial 25, we only count until we get to the occurrence of code 21, three transitions away. Similarly, counting from the second occurrence of code 25, we avoid a repetition of code 29 by only counting until we reach code 13, two transitions away. Thus, in the example sequence given above, we find that code 25 is one transition from code 13, then find that code 25 is two transitions from code 5 and three transitions from code 21. Next we count from code 13 (the second code in the sequence) and find that code 13 is one transition from code 5, two transitions from code 21 etc.

Notice that three estimates of the distance between code 25 and code 13 are obtained, since we count from the first example of code 25, then from the second example of code 25, and finally from the first example of code 13. All three estimates are averaged to get the mean number of transitions between code 25 and code 13. Note also, however, that this counting scheme gives no estimate of the distance between code 25 and code 9. This is because when counting from the first example of code 25, we see that code 25 is repeated before getting to code 9. Counting from the second 25 in the sequence, code 29 is repeated before code 9 is encountered. When the counting scheme does not give any estimates of the distance between two codes, a distance estimate equal to the number of codes in the codebook is used, effectively giving a maximum estimate of the distance.

Now that the method used to estimate the distances between isocode regions has been discussed, we can explain why code sequences were generated by random walks. Suppose we are trying to estimate the distance between isocode region 26 and region 22 from the sequence of VQ codes. If the tongue makes a relatively smooth downward motion from region 26 to region 22, we expect to see the VQ code sequence: 26, 10, 22. Notice that this code sequence gives a good estimate of the number of regions between region 26 and region 22. In contrast, if the tongue takes a random walk, it is fairly likely to travel to code 18, 23, 4, 28, 12 or even code 13 for that matter, before it gets to code 22. Typically, a random path is a longer than necessary way to get from one point to another. It is only by averaging the information from such random paths that we get distance estimates that should be monotonically related to actual distance estimates. Presumably the continuity mapping algorithm will work better given smoother tongue motions, as long as the tongue motions still travel through each of the isocode regions.

The relative positions of the isocode regions were estimated from the average intercode transition distances using non-metric multidimensional scaling (MDS). Multidimensional scaling calculates relative point positions from interpoint distances by starting with some initial configuration of points in space, and then moving the points until the distances between the points are nearly monotonically related to the desired interpoint distances (Dillon and Goldstein, 1984, provide more information about MDS). The MDS algorithm moves the points using gradient descent on an error measure, *stress*, which is a measure of the departure from a monotonic relationship between the interpoint distance as determined by MDS and desired interpoint distances.

While MDS is capable of producing solutions with different numbers of dimensions, the interpoint distances were generated from two-dimensional data, so solutions of more than two dimensions will only be fitting the noise in the intercode distance estimates. Because we know that the correct MDS solution is two-dimensional (i.e. the articulator map is two-dimensional), all the continuity maps have two dimensions.

The gradient descent minimizations performed by MDS can find local minima as well as global minima, where local minima are solutions that minimize stress in a local region, but which are not the best solution. The best way to avoid using a solution that is merely a local minimum is to run the MDS algorithm from a variety of different random starting configurations. So, to avoid local minima, five two-dimensional solutions were found for each set of interpoint distances, using a different initial configuration of points to get each solution. Since there were 21 random walks, 105 different solutions were found. For each random walk, the solution with the lowest stress value was used for further analysis, giving one best solution for each random walk. The resulting maps are called continuity maps (CMs) because they are made based on the fact that articulators move in a continuous fashion.

To show that different random walks lead to very similar CMs, CMs made from different random walks were compared using *generalized Procrustes analysis* (Gower, 1975), a technique for rotating, translating, reflecting, and scaling (only uniform scaling is allowed) configurations to make them maximally similar, and then calculating a measure of how similar the different configurations are. Figure 3 illustrates Procrustes analysis (Lederman, 1984), the basic component of generalized Procrustes analysis. Two configurations of three points each are shown in Figure 3A.

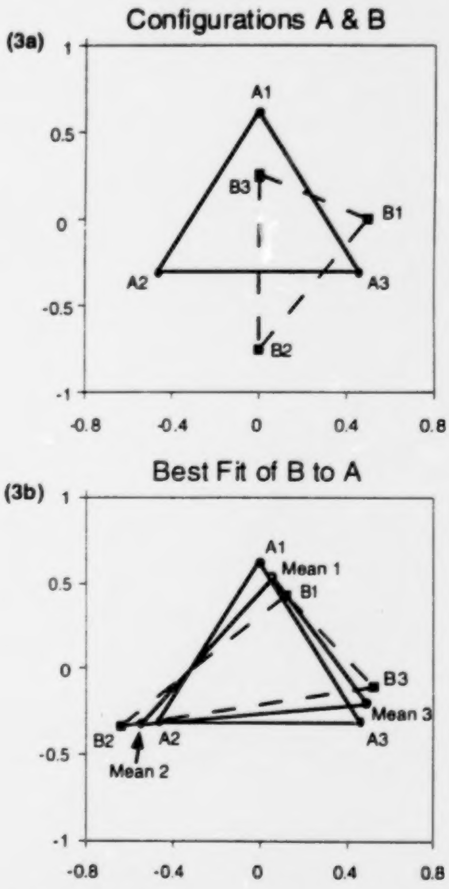


Figure 3. Example of Procrustes analysis. Since only the relative configuration of points is of interest, the axes are not relevant and therefore are unlabeled. Plot 3A shows the two configurations that need to be compared. In plot 3B, configuration B has been rotated to best fit configuration A, and the mean configuration is shown.

As you can see, the configurations are not aligned and would not be identical even if they were better aligned. Figure 3B shows the result of using Procrustes analysis to rotate, reflect, scale, and translate configuration B to best fit configuration A. In a perfect fit, point A1 would be directly over point B1, A2 would be directly over B2, and A3 would be directly over B3, which is not the case for these two configurations. To calculate the deviation from a perfect fit, the configurations are compared to the mean configuration, also shown in Figure 3B, by finding the square root of the mean squared distance between each point and the corresponding mean position. The mean configuration can also be used as the estimate of the true configuration. For the extension of this procedure to more than two configurations (the extension is called *generalized* Procrustes analysis), refer to Gower (1975).

The results of generalized Procrustes analyses of the CMs generated by random walks of the same length are shown in Figure 4. The error in Figure 4 is given in the same units that are used on the axes of Figure 5A. These errors are extremely small—for example, when there are at least 10 repetitions of each code, an error bar representing the standard deviation between a point in a CM and the corresponding point in the mean CM would be approximately the size of the characters used to label the codes in Figure 5A. Clearly, by the time there have been fifty repetitions of each code, CMs generated from different random walks are nearly identical.

4. EVALUATING THE CONTINUITY MAP

The crucial comparison to be made is between the relative positions of the codes shown in the CM in Figure 5A and the corresponding positions in the known AM shown in Figure 5B. The position of a code, code 7 for example, in Figure 5B is the mean of all the tongue positions (from the articulator map in Figure 2) that produced a sound encoded as 7. The CM has been rotated and scaled to best fit the mean tongue positions, but the relative positions of the codes in the CM were not changed.

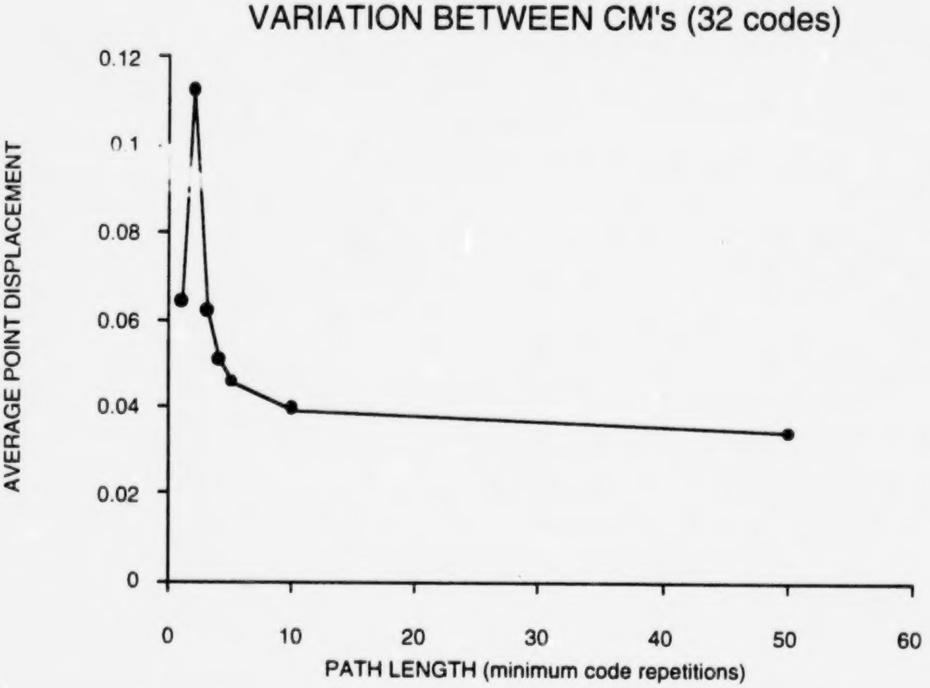


Figure 4. This plot shows the average distance between the points in the continuity maps and their corresponding mean positions as a function of the minimum number of repetitions of each code in the path. A small average distance indicates that continuity maps made from different random walks are similar.

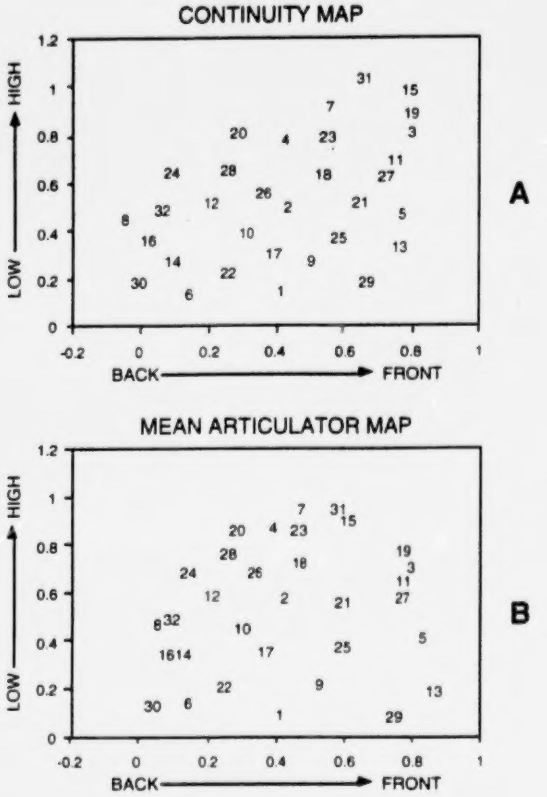


Figure 5.A) The continuity map(CM) showing estimated tongue positions. The CM has been rotated, reflected, scaled, and translated to best fit the mean articulator map shown in 5B, but the relative positions of the points in the continuity map have not been changed. *B)* The means of all the tongue positions that give rise to each code.

While the CM does show signs of non-uniform stretching relative to the plot of mean tongue positions, the relative positions of the codes are clearly similar in the two plots. The stretching can be attributed mostly to the thinness of the isocode regions when the tongue is extremely far forward. Since each isocode region is one transition away from its nearest neighbors, MDS tries to make the distances between neighboring isocode regions approximately equal. This means that the distance between neighboring large isocode regions should be about the same as between neighboring small isocode regions. Thus, the thin isocode regions that occur when the tongue is fronted are represented as taking up relatively larger regions in the CM than in the AM, distorting the CM relative to the AM.

Despite the distortions, the x-axis of the CM correlates well with the fronting axis of Figure 5B as seen in Figure 6A, which plots the position of the codes on the x-axis of the CM versus the fronting axis of Figure 5B. The rank-order correlation between the positions is 0.98, showing that the position of a code in the CM can give us information about the relative fronting of the tongue. Similarly, the y-axis of the CM is compared to the height axis of Figure 5B in Figure 6B. The rank order correlation between the height given by the CM and the actual height is 0.97.

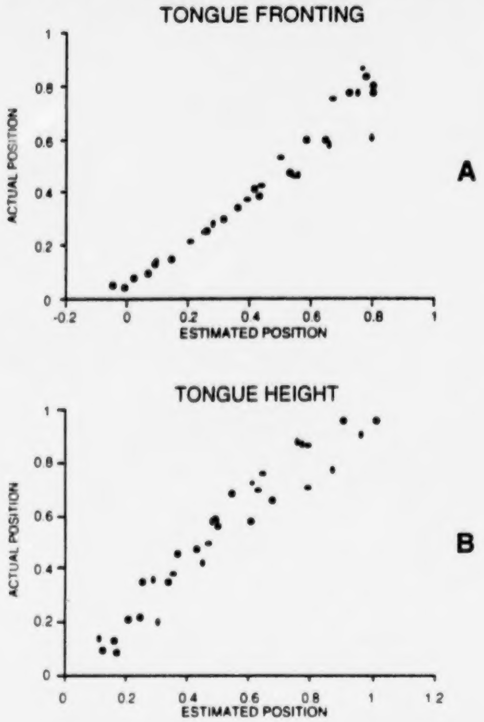


Figure 6.A) Each point represents a single code; the x-axis is the position of the code on the fronting axis of the continuity map and the y-axis shows the position of the code on the fronting axis of the mean articulator map. *B)* The continuity map height positions are compared to the mean articulator map height positions.

Similar results were found for three different codebooks with 64 codes and one codebook with 128 codes. Both the 64- and 128-code books had disjoint isocode regions when the tongue was low and very far back, in nearly the same areas as in the 32-code book. In the cases where disjoint isocode regions did occur, they were still relatively close together. As before, the correlations between estimated and actual articulator positions were high (0.95 or above, median correlation = 0.98) and the particular random walk used to make the continuity maps made very little difference as long as the random walk contained at least 10 repetitions of each code.

5. DISCUSSION

All thirty of the continuity maps created from random walks with at least 10 repetitions of each code (this includes those with at least 50 repetitions of each code) gave good estimates of the relative locations of the mean articulator positions. The high correlations between the continuity maps and the average tongue positions clearly show that the continuity maps can be used to estimate the relative locations of the mean tongue positions for this synthesized data set. Of course, the ability of the continuity maps to represent the relative tongue positions also depends on how well the centroids of the isocode regions approximate the actual tongue positions. Since the ability to estimate the mean tongue positions stays approximately constant as the number of codes increases, but the mean tongue positions become better estimates of the actual tongue positions, the accuracy of the estimates of relative tongue positions increases as the number of codes increases.

The consistently high correlations found with different VQ codebooks were surprising because, although the positions of codes in the continuity maps should be topologically similar to the positions of the centers of gravity of the isocode regions, the positions can be uncorrelated even if the two maps are topologically identical. Non-uniform stretching of one map relative to the other can decrease the correlation while maintaining topological similarity. In this study, some non-uniform stretching was found, particularly for front tongue positions, but the effect on the overall relative positions was small. The non-uniform stretching may be more prominent in continuity maps of natural speech.

Once a continuity map (CM) is created from training data, it can be used to give relative articulator position estimates for subsequent speech, without the algorithm ever getting any information about the absolute positions of the articulators. One possible use for the continuity mapping technique would be training the deaf to speak. For example, the algorithm could be used to create a continuity map from recordings of an instructor's voice. Once the continuity map is made, new speech sounds made by the instructor could be vector quantized, and the position of the vector quantization code in the CM could be used as an estimate of the instructor's articulator configuration. The instructor's articulation could then be displayed on a computer screen for the students to imitate. While only the relative positions are recovered from the technique described here, the absolute positions of the articulators can presumably be determined from only a few examples of acoustic signals created from known articulator positions, because only rotation and scaling information is needed to get the absolute positions from the relative positions.

A weakness of continuity mapping is that it only uses information from one short-time window of speech to determine articulator positions. This will make the technique less robust under noisy conditions. By treating the CM as a hidden Markov model (Huang et al. 1990), it should be possible to use information from several windows of speech. One way to do so would be to treat the VQ codes as hidden Markov model states, then estimate transition probabilities between each of the codes in the CM and find the probability distributions of the observed acoustic vectors around the VQ reference vectors (the prototype vectors used in the nearest neighbor categorization). After making these extensions, it should be possible to calculate the path through the CM with the highest probability of creating an observed acoustic sequence. Research in this direction will have to address the computational problems of learning the transition probabilities for such a large network (normal hidden Markov models have many fewer possible transitions).

There are two main conclusions to draw from these results. The first is that, even though the data set contained a few cases where different articulator positions created the same derived acoustic parameters, there was enough information in the data set to find a rough mapping from acoustic information to the simulated articulator positions. If this were not the case, the continuity mapping procedure could not have found the mapping. The second conclusion is about the technique itself: using only unsupervised learning, the continuity mapping technique was able to recover information about the positions of moving objects. This suggests that continuity mapping may have applications beyond speech (Hogden et al., 1992b give an example), since objects in the world move continuously and we often need to obtain knowledge about their physical positions from sensory information.

REFERENCES

Atal, B. S. (1975). Towards determining articulator positions from the speech signal. In G. Fant (Ed.), *Proceedings of the 1974 Stockholm Speech Communication Seminar* (pp. 1-9). New York: Wiley.

Atal, B. S., Chang, J. J., Mathews, M. V., & Tukey, J. W. (1978). Inversion of articulatory-to-acoustic transformation in the vocal tract by a computer-sorting technique. *Journal of the Acoustical Society of America, 63(5)*, 1535-1555.

Boe, L. J., Perrier, P., & Bailly, G. (1992). The geometric vocal tract variables controlled for vowel production: Proposals for constraining acoustic-to-articulatory conversion. *Journal of Phonetics, 20,* 27-38.

Coker, C. (1976). A model of articulatory dynamics and control. *Proceedings of the IEEE, 64(4),* 452-460.

Dillon, W., & Goldstein, M. (1984). *Multivariate analysis: Methods and applications.* New York: John Wiley & Sons.

Fant, G. (1970). *Acoustic theory of speech production* (2nd ed.). The Hague: Mouton & Co.

Flanagan, J. (1972). *Speech analysis, synthesis, and perception* (2nd ed.). New York: Springer-Verlag.

Fowler, C., & Turvey, M. (1980). Immediate compensation in bite-block speech. *Phonetica, 37,* 306-326.

Gower, J. (1975). Generalized Procrustes analysis. *Psychometrika, 40(1),* 33-51.

Gray, R. (1984). Vector quantization. *IEEE Acoustics, Speech, and Signal Processing Magazine,* 4-29.

Harshman, R., Ladefoged, P., & Goldstein, L. (1977). Factor analysis of tongue shapes. *Journal of the Acoustical Society of America, 62(3),* 693-707.

Hogden (1991) *Low-dimensional phoneme mapping using a continuity constraint.* Unpublished doctoral dissertation, Stanford University.

Hogden, J., Löfqvist, A., Gracco, V., Oshima, K., Rubin, P., & Saltzman, E. (1993). Inferring articulator positions from acoustics: an electromagnetic midsagittal articulometer experiment. *Journal of the Acoustical Society of America, 94(3),* 1764(A).

Hogden, J., Rubin, P., & Saltzman, E. (1992a). An unsupervised method for learning to track tongue position from an acoustic signal. *Journal of the Acoustical Society of America, 91(4),* 2443. (A)

Hogden, J., Saltzman, E., & Rubin, P. (1992b). Unsupervised neural networks that use a continuity constraint to track articulators. *Journal of the Acoustical Society of America, 92(4),* 2477. (A)

Huang, X. D., Ariki, Y., & Jack, M. (1990). Hidden Markov models for speech recognition. Edinburgh: Edinburgh University Press.

Jordan, M., & Rumelhart, D. (1992). Forward models: supervised learning with a distal teacher. *Cognitive Science, 16,* 307-354.

Kawato, M. (1989). Motor theory of speech perception revisited from minimum torque-change neural network model. In *8th Symposium on Future Electron Devices* (pp. 141-150).

Kohonen, T. (1988). The neural phonetic typewriter. *Computer,* 11-22.

Kuc, R., Tutuer, F., & Vaisnys, J. R. (1985). Determining vocal tract shape by applying dynamic constraints. In Proceedings of the International Conference on Acoustics Speech & Signal Processing, 1101-1104, New York: IEEE.

Ladefoged, P., Harshman, R., Goldstein, L., & Rice, L. (1978). Generating vocal tract shapes from formant frequencies. *Journal of the Acoustical Society of America, 64(4),* 1027-1035.

Lederman (1984). Orthogonal Procrustes Analysis. In E. Lloyd (Ed.), *Handbook of applicable mathematics* (pp. 761-781). New York: John Wiley & Sons.

Liberman, A., Cooper, F., Shankweiler, D., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review, 74(6),* 431-461.

Liberman, A., & Mattingly, I. (1985). The motor theory of speech perception revised. *Cognition, 21,* 1-36.

Lindblom, B., Lubker, J., & Gay, T. (1979). Formant frequencies of some fixed-mandible vowels and a model of speech motor programming by predictive simulation. *Journal of Phonetics, 7,* 146-161.

Maeda, S. (1979). An articulatory model of the tongue based on a statistical analysis. *Journal of the Acoustical Society of America, 65,* S22.

Maeda, S. (1989). Compensatory articulation during speech: evidence from the analysis and synthesis of vocal-tract shapes using an articulatory model. In NATO AI Series, Kluwer Academic Publishers.

Markel, J., & Gray, A. (1976). Linear prediction of speech. New York: Springer-Verlag.

McGowan, R. (1994). Recovering articulator trajectories using task dynamics and a genetic algorithm. *Speech Communication, 14,* 19-48.

Mermelstein, P. (1973). Articulatory model for the study of speech production. *Journal of the Acoustical Society of America, 53(4),* 1070-1082.

Morrish, K., Stone, M., Shawker, T., & Sonies, B. (1985). Distinguishability of tongue shape during vowel production. *Journal of Phonetics, 13,* 189-203.

O'Shaughnessy, D. (1987). Speech Communication: Human and Machine. New York: Addison-Wesley.

Papcun, G., Hotchberg, J., Thomas, T., Laroche, F., Zacks, J., & Levy, S. (1992). Inferring articulation and recognizing gestures from acoustics with a neural network trained on X-ray microbeam data. *Journal of the Acoustical Society of America, 92(2),* 688-700.

Rahim, M. G., Kleijn, W. B., Schroeter, J., & Goodyear, C. C. (1991). Acoustic to articulatory parameter mapping using an assembly of neural networks. Proceedings of the International Conference on Acoustics, Speech, and Signal Processing, 485-488.

Rubin, P., Baer, T., & Mermelstein, P. (1981). An articulatory synthesizer for perceptual research. *Journal of the Acoustical Society of America, 70(2),* 321-328.

Schroeter, J., Meyer, P., & Parthasarathy, S. (1990). Evaluation of improved articulatory codebooks and codebook access distance measures. Proceedings of the IEEE International Conference on Acoustics, Speech, & Signal Processing, 393-396.

Schroeter, J., & Sondhi, M. (1992). Speech coding based on physiological models of speech production. In S. Furui & M. Sondhi (Eds.), *Advances in speech signal processing* (pp. 231-267). New York: Marcel Dekker, Inc.

Shirai, K., & Kobayashi, T. (1986). Estimating articulatory motion from speech wave. *Speech Communication, 5,* 159-170.

Smith, N. (1973). The acquisition of phonology: A case study. Cambridge: Cambridge University Press.

Sondhi, M. (1979). Estimation of vocal tract areas: the need for acoustical measurements. IEEE Trans. *ASSP, 27(3),* 268-273.

Stevens, K., & House, A. (1955). Development of a quantitative description of vowel articulation. *Journal of the Acoustical Society of America, 27*(3), 484-493.

FOOTNOTES

[*] Appears in *Bulletin Communication Parlee* (1994).

[†] Institute for Mathematical Behavioral Sciences. University of California at Irvine.

[‡] Also Yale School of Medicine, Department of Surgery.

Haskins Laboratories Status Report on Speech Research
1993, SR-115/116, 45-55

Prosodic Patterns in the Coordination of Vowel and Consonant Gestures*

Caroline L. Smith†

That vowels and consonants can appear independent of one another has been suggested by both autosegmental representations (e.g., Archangeli, 1985; McCarthy, 1981, 1982, 1989; Prince, 1987; Steriade, 1986) and analyses of articulatory and acoustic data (e.g., Fowler, 1980, 1981, 1983; Joos, 1948; Öhman, 1966, 1967). In phenomena such as vowel harmony, phonological processes involving vowels sometimes operate as if an intervocalic consonant were absent; likewise, consonantal harmony processes or co-occurrence restrictions may ignore intervening vowels. In speech production, it has been suggested that there are distinct, identifiable vowels and consonants, but that the two classes of sounds may be produced at least partially simultaneously. It is hypothesized that these periods of simultaneous production ("coproduction," as discussed in Fowler, 1980) are responsible for the context-dependent influences of vowels on consonants, and vice versa.

This study focuses on the organization of the temporal relations between vowels and consonants. It is proposed that languages may choose among alternative ways of coordinating vowels and consonants, and that these alternatives underlie differing prosodic properties that languages exhibit, such as the timing patterns traditionally described as stress-, syllable- or mora-timing. The approach taken here involves defining consonants and vowels in terms of articulatory gestures, following Browman and Goldstein's Articulatory Phonology (e.g., Browman & Goldstein, 1986, 1992). This framework provides a phonological description that explicitly specifies how consonants and vowels are produced, which makes it possible to predict how differences in articulatory coordination could result in different prosodic characteristics.

In Articulatory Phonology, a gesture is both a primitive of phonological representation and an abstract, dynamic unit of action that controls the coordinated movement of one or more articulators. The spatial and temporal properties of each gesture are specified in terms of tract variables (Saltzman, 1986; Saltzman & Kelso, 1987; Saltzman & Munhall, 1989). These are variables such as Lip Aperture or Tongue Body Constriction Degree, which are characterized by categorically-valued descriptors. These descriptors specify the parameters of the task(s) involved in producing the gesture, a task typically being the formation of a constriction in some part of the vocal tract. For instance, in a bilabial gesture such as for /p/, the goal, specified using the Lip Aperture tract variable, is to close the lips together. Temporally, the gesture is specified by the parameters of the tract variable(s) that determine the time course of the movements associated with that gesture (Browman & Goldstein, 1986, 1990 a,b, 1992; Saltzman, 1986; Saltzman & Kelso, 1987; Saltzman & Munhall, 1989).

In Articulatory Phonology, gestures are abstract, phonological units; by positing a relation between tract variables and actual articulators, the gestures can be associated with the measurable movements of the articulators for the different vowels and consonants. In this way the

---

This work was supported by NSF grant BNS 8820099 and NIH grant DC 00121 to Haskins Laboratories. Support during the preparation of this manuscript was provided by a fellowship from the Fondation Fyssen. I thank Cathe Browman, Louis Goldstein, Ian Maddieson, and Ignatius Mattingly for advice and comments on various versions of this material, but do not wish to imply that any of them is responsible for remaining flaws.

temporal characteristics of a gesture can be estimated from movement in the part of the vocal tract that would be controlled by that gesture. Of course, measuring the movements of just one articulator is at best an approximation, because a gesture typically involves multiple articulators. For instance, a bilabial closing gesture involves one tract variable (Lip Aperture), but several articulators—the jaw, lower and upper lips.

Because duration is an intrinsic property of gestures, they are well-suited to serve as the units of representation in processes that crucially involve the temporal extent of phonological properties. (Steriade's 1990 gestural analysis of Dorsey's Law provides an example of such a process.) In order to use gestures to represent such processes, it is desirable to specify a structure for the temporal relations among the gestures. It would be expected that only a limited set of the possible temporal relations among gestures would be stable in a given language (Goldstein, 1989). Because gestures are defined in terms of possible vocal tract goals, they are inherently constrained by the physical possibilities of the vocal tract. Their representation is additionally constrained in Articulatory Phonology by limits on the permissible values for dynamic parameters, and can be further constrained by a specific model of the temporal coordination between vowels and consonants.

1 MODELS

Two models of this coordination will be compared here. Both models have been proposed as representations of consonant-vowel relations in English, although they also apply to other languages as well (or better).

In one of these models, each oral gesture is coordinated with the oral gesture preceding it and the one following it (Browman & Goldstein, 1990a); thus, consonants and vowels are mutually coordinated. Vowels can be coordinated with respect to consonants, and vice versa, rather than being exclusively coordinated with other vowels. This arrangement means that the temporal properties of consonants can affect vowels, either as a result of the properties of individual consonants or as a function of the number of consonants. How such an effect comes about depends on exactly how the consonants and vowels are coordinated. For example, in a VCV sequence, suppose one phase of the consonant's production (e.g., the achievement of closure) is coordinated with the preceding vowel and that the following vowel is coordinated with respect to a later phase in the consonant (e.g., its release). If the duration of the consonant were to increase, the time between the two phases of the consonant, and hence between the vowels, might be expected to increase. This model, illustrated on the left-hand side of Figure 1, will be referred to as "combined vowel-and-consonant timing".

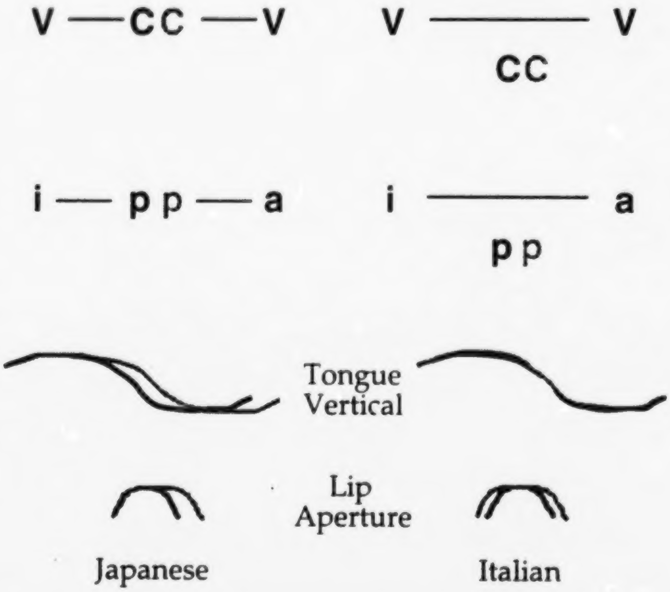


Figure 1. Two possible models of coordination between consonant and vowel gestures: The combined vowel-and-consonant model on the left, and the vowel-to-vowel model on the right.

Another model gives special status to the vowels. Öhman's (1967) analysis of vowel-to-vowel coarticulation proposed that the production of individual consonants is superimposed on the continuous production of vowels. Similarly, Fowler (1983) summarized a variety of experimental and phonological evidence suggesting that, at least for a sequence of stressed monosyllables, vowels are produced continuously and consonants are coordinated with them. That is, the consonants are produced separately from vowels but organized temporally with respect to them; since the production of vowels is continuous, consonants will overlap them (Fowler, 1983).

One problem in comparing the model suggested by Fowler with the combined vowel-and-consonant model is that the two models are based on somewhat different concepts of vowel production. If the vowels and consonants are both defined in terms of discrete articulatory gestures, as in Articulatory Phonology, then the production of vowels is not a strictly continuous cycle as it is in Fowler's model. The kind of vowel-based model assumed here is similar to Fowler's, in that it retains the independence of vowels from consonants that characterizes Fowler's model, but it treats each vowel as a discrete gesture that must be coordinated with other vowel gestures. In this model, the oral gesture for each vowel is coordinated with the oral gesture for the preceding vowel, and each consonantal oral gesture is also coordinated with the oral gesture for the preceding vowel.

In this kind of model, consonants are essentially irrelevant to the temporal organization of the vowels, which is dependent on the foot structure (patterning of the stressed and unstressed units). The production of vowels should not be affected by the number of consonants or by any other property of them, such as inherent differences in duration among types of consonants. This prediction contrasts with the prediction of the combined vowel-and-consonant model that the differences in the consonant(s) would affect the vowels. The model of this type similar to Fowler's is shown on the right-hand side of Figure 1, and will be referred to as "vowel-to-vowel timing".

The distinction between these two possible ways of coordinating vowels and consonants can be seen when the duration of the intervocalic consonant changes. The lower half of Figure 1 compares the predictions of the two models for utterances with single or geminate consonants. Stylized traces of tongue and lip movements for the sequence /ipa/ are shown by black lines. When the intervocalic consonant is a geminate /p/ (as shown by the light gray lines in the figure) rather than a singleton, the combined vowel-and-consonant model predicts a reorganization in the timing of the movements associated with the vowels. As the left-hand side of Figure 1 suggests, this might be a delay in the second vowel until after the longer consonant has been produced, illustrated by the Lip Aperture trace showing a longer period of closure. But, as shown on the right, in vowel-to-vowel timing where the coordination of the vowels is independent of the consonant, the prediction would be that no change would occur in the movements associated with the vowels.

These two models reflect, at the very least, different logical possibilities for coordinating consonant and vowel events. It is hypothesized here that both do occur, but that they are found in languages with different prosodic structures, with the vowel-to-vowel model of organization underlying languages whose rhythm has been described as being based on vowels ("stress- or syllable-timed"), and the combined vowel-and-consonant model underlying languages that have been described as "mora-timed". The best-known example of a mora-timed language is Japanese (Vance, 1987), in which both vowels and coda consonants "count" in determining the number of moras. Examples of mora-counting in Japanese are given in Table 1.

Conversely, for languages in which the rhythm is determined primarily on the basis of the vowels, vowel-to-vowel timing seems more appropriate. Italian will be used here as an example of this kind of language. There is considerable debate over whether Italian is best classified as stress-timed or as syllable-timed (see e.g., Dauer, 1983; Bertinetto, 1983); what is relevant (and widely accepted) is that its rhythm is centered around the vowels, thus, it is in a different category from Japanese. In this study it is the two models of temporal organization of gestures that are being compared, but it is also suggested that these models provide a way of understanding differences between languages that fall into the different rhythm categories known as syllable- and mora-timing.

Table 1. *Number of moras in Japanese words.*

"a"	1 mora	"kan"	2 moras
"ka"	1 mora	"kana"	2 moras
"an"	2 moras	"kanna"	3 moras

These two models of temporal organization have been formulated here in terms of an abstract level of gestural control, but to test whether they accurately model consonant-vowel relations in Japanese and Italian requires data showing the actual movements of the articulators that are associated with these gestures. In this study, articulatory data were collected that make it possible to measure the continuous movements associated with the vowel and consonant gestures and examine their temporal behavior. By using articulatory data, it is possible to separate movements associated with consonants, for example lip movement for bilabials, from movements of the tongue associated with vowels. These gestures cannot be measured separately in acoustic data where there is only one channel of data that incorporates both vowels and consonants.

2 EXPERIMENTAL METHOD

In this experiment, such movements were measured using the NIH X-ray microbeam facility at the University of Wisconsin, Madison (Nadler, Abbs, & Fujimura, 1987; Westbury, 1991). This system recorded the movements of the articulators by means of a microscopic X-ray beam that tracks tiny gold pellets attached to the midline of the tongue, the upper and lower lips, and the lower incisor (to measure jaw movement).

Figure 2 shows a midsagittal cross-section of the vocal tract indicating the approximate positions of the pellets on the articulators for the speakers in this experiment.

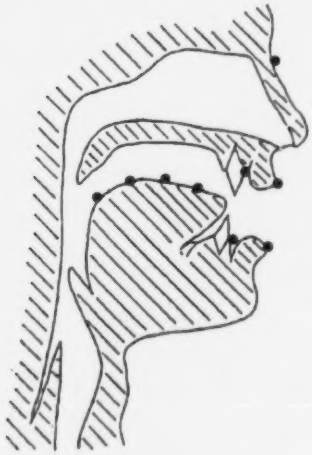


Figure 2. **Mid-sagittal cross-section of the vocal tract, showing approximate positions of the microbeam pellets on the speaker's articulators (after Abbs & Nadler, 1987).**

Microbeam data consist of trajectories of the pellets in a two-dimensional coordinate system over time, thus showing the movements of the articulators.

Three speakers each of Japanese and Italian produced disyllabic nonsense words consisting of the vowels /a/ and /i/ with single and geminate intervocalic consonants. The target utterances are shown in Table 2. These utterances were produced within carrier phrases designed to provide phonetically similar contexts in the two languages: "Boku wa _____ mo aru" ("I have a _____, too.") for Japanese, and "Dica _____ molto" ("Say _____ again and again") in Italian.

Table 2. *Target utterances.*

map(p)i	mip(p)a
mat(t)i	mit(t)a
mam(m)i	mim(m)a
man(n)i	min(n)a

Note that utterances of the form /matti/ were not collected for Japanese because a rule of palatalization changes the [t] to [tʃ] before [i]. Experimental constraints resulted in considerable variation in the number of tokens per utterance produced by the speakers: for the Japanese speakers (J1-J3), there were typically 15 to 20, for Italian speaker 1, about 10, and for the other two Italian speakers (designated I2 and I3), about 5.

The positions of the microbeam pellets over time were recorded, and traces of their horizontal and vertical movements were used to measure the intervals between various significant events. The pellet traces chosen for measurement were those that could be most directly associated with tract variables involved in the production of the vowel and consonant gestures. In this way a connection could be made between significant events in the articulatory movements and in the gestures. For most speakers, four pellets were attached to the tongue, as in Figure 2, and the horizontal movement of the rearmost pellet and vertical movement of next-rearmost were associated with the vowel gestures.[1] Consonantal gestures for bilabials were associated with the Lip Aperture trace, calculated as the vertical distance between the upper and lower lip pellets.

The locations of the measured events were determined algorithmically. For example, movement onset was identified as the time when the velocity of the movement exceeded zero by a predetermined threshold, and the achievement of target of a gesture was associated with the time at which the velocity of the movement slowed down to no greater than the threshold value, and approached a displacement plateau. For the tongue movement traces, this threshold was ±10% of the most extreme velocity recorded for a particular trace for a given speaker.

3 RESULTS

The different timing patterns that were observed in these utterances will be illustrated by individual tokens chosen to be representative of statistically significant effects.[2] For details of the statistical analyses, see Smith (1992).

In Figure 3, Japanese speaker J1's production of the utterance /mipa/ is illustrated by the lower of the two acoustic waveforms and the solid black lines in the articulatory movement traces. The vertical lines in the movement traces indicate the times at which the tongue approached the target locations for the vowel gestures. To test the hypothesis that Japanese shows re-organization of the timing between the vowel gestures when the duration of the consonant increases, this utterance was compared to the corresponding utterance with a geminate /p/, shown in Figure 3 in the top acoustic waveform and the light gray lines in the movement traces.

The two utterances are lined up in this figure at the offset of the initial /m/ in the Lip Aperture movement trace. Notice that in the utterance with a geminate intervocalic consonant, the second vowel reaches its target position later relative to the first vowel, and the plateau region for the first vowel is longer preceding the geminate. These differences between the utterances with single and geminate consonants were statistically significant for all Japanese speakers. The change in the relation between the vowels that is observed when the consonant is a geminate suggests that the combined vowel-and-consonant timing model is appropriate for Japanese.

The contrast between single and geminate consonants shows up in a quite different way in Italian. Productions of /mipa/ and /mippa/ by Italian speaker I1 are shown in Figure 4. For this speaker, when the intervocalic consonant is a geminate, the movements of the tongue associated with the vowels remain essentially unchanged from the utterance with a single intervocalic consonant.

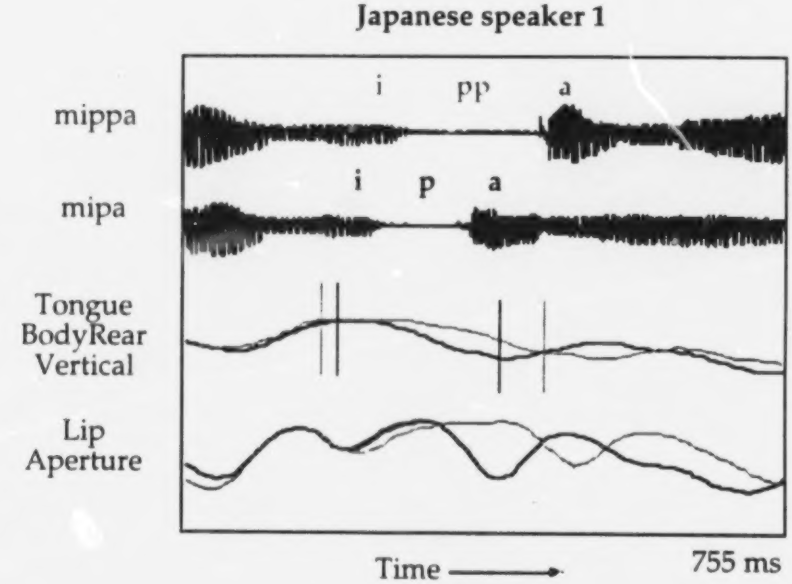


Figure 3. Productions of /mipa/ (dark lines) and /mippa/ (light lines) by Japanese speaker J1. The vertical lines mark the times of achievement of target in the tongue movement associated with the vowel gestures.

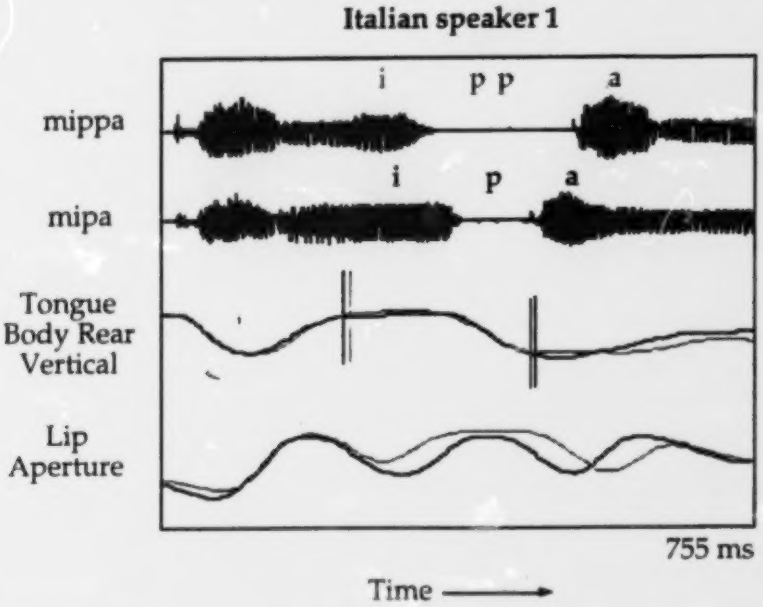


Figure 4. **Productions of /mipa/ (dark lines) and /mippa/ (light lines) by Italian speaker I1. The vertical lines mark the times of achievement of target in the tongue movement associated with the vowel gestures.**

In particular, the times at which the two vowels reach their target positions (shown by the vertical lines) are virtually the same in the utterances with single and geminate consonants. This result suggests that the consonant has not affected the relative timing of the vowels. For this speaker, there was no significant difference in the interval between vowel targets between utterances with single and geminate consonants. As predicted by the vowel-to-vowel timing model, the vowels seem to be coordinated independently of the consonant.

For speaker I3, shown in Figure 5, and also speaker I2, there was a small, statistically significant shortening[3] of the duration of the interval between vowel targets with an intervening geminate consonant. Whereas in Japanese the interval between vowel targets was longer with a geminate intervocalic consonant, for Italian speakers I2 and I3 this interval was slightly shorter. Although these two Italian speakers showed some changes in vocalic durations when the intervocalic consonant was longer, they resembled speaker I1 in that the total duration of the articulatory movements from start to finish of the utterances were extremely similar regardless of the length of the intervocalic consonant. For speakers I2 and I3 there seems to be a trade-off in the durations of the interval from the target to the offset of the first vowel and the interval from the offset of the first vowel to the target of the second vowel. Note that in the utterance with the geminate consonant, the plateau region for the first vowel is shorter but the interval from the offset of the first vowel to the target of the second vowel is much longer, with the result that the time from vowel target to vowel target is not very different from that observed with the single consonant.

This pattern suggests that since the tendency for speakers I2 and I3 is that the target-to-target interval does not vary much, all the Italian speakers are more alike than may initially appear to be the case. Nonetheless, the fact that speakers I2 and I3 do show a difference between the single and geminate utterances for the interval between vowel targets, implies that the length of the consonant may affect the vowels to some extent. Therefore, speakers I2 and I3, who behaved similarly, contradict the strongest form of the vowel-based hypothesis, which predicted that the vowels would be unaffected by a change in the consonants. The strong form of the vowel-to-vowel timing hypothesis is completely borne out only for the Italian speaker I1. However, all the Italian speakers clearly differ from the Japanese speakers. It remains unclear, however, to what extent the patterns shown by the Italian speakers can be described by the vowel-based hypothesis alone.

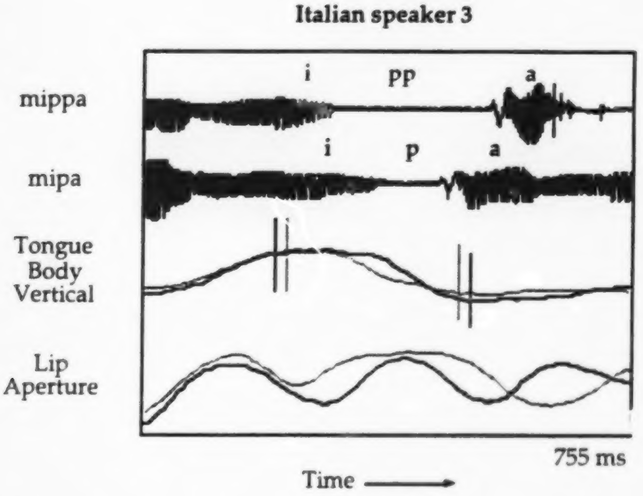


Figure 5. **Productions of /mipa/ (dark lines) and /mippa/ (light lines) by Italian speaker I3. The vertical lines mark the times of achievement of target in the tongue movement associated with the vowel gestures.**

4 MODELING

To illustrate how the measured durational changes between utterances with single and geminate consonants could arise from limited changes to the relations among gestures, models of timing were constructed for Japanese and Italian, with the timing relations among the gestures specified in terms of phasing. The modeling was done by manipulating parameters that specify the temporal characteristics of the individual vowel and consonant gestures and their relative phasing. For each speaker, these parameters were manipulated to create the best possible model structured in accordance with the hypothesized timing organization; that is, the models for Japanese speakers used the vowel-and-consonant organization and the models for Italian speakers the vowel-to-vowel based organization. The basic structures for the models are shown in Figure 6, where the rounded boxes represent consonant and vowel gestures and the lines between them show the phasing relations that were specified in the models.

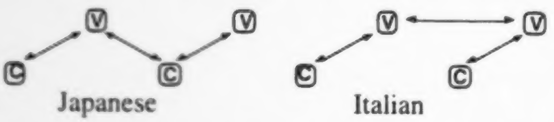


Figure 6. **Intergestural phasing relations used in the models.**

Figure 7 illustrates how the differences between utterances with single and geminate consonants were modeled; the boxes represent abstract gestures and the lines between them connect phases in the gestures that were specified to occur at the same times. The black lines correspond to the model for the utterance with a single consonant, and the light gray lines show the model for the utterances with a geminate consonant.

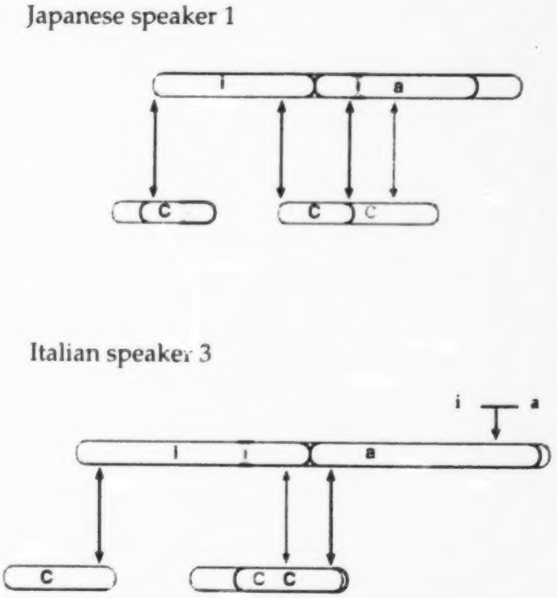


Figure 7. **Structures of the models for Japanese and Italian. Each box corresponds to a gesture: the width is scaled to the mean duration of the articulatory movements associated with that gesture, from onset to end of the period of activation, for the individual speaker.**

These diagrams are drawn to scale, and the width of the boxes reflect the measured durations, for one speaker of each language (J1 and I3), of the movements associated with the different gestures.

These models were used to predict durations of intervals between pairs of events that had been identified in the articulatory movements, such as the events marked by vertical lines in Figures 3, 4, and 5. The differences in these durations between the utterances with single and geminate consonants were then calculated, and the predicted and measured differences were compared. The modeling procedure minimized the number of parameters whose values varied between the utterances with single and geminate consonants, while optimizing the r^2 correlation between the measured and predicted differences in durations.

For each speaker, one model was optimized using the structure hypothesized for that speaker's language. The same model was used for Italian speakers I2 and I3. In addition, all of these models were tested on the data for every speaker, so that both structures were tested on speakers of both languages. Speakers of each language were best modeled (that is, the fit had the highest r^2) by the models using the structure that had been hypothesized to reflect the patterns of their language. Even when using models optimized for other speakers, better fits were found with models for speakers of the same language, which lends support to the initial assumption that the two languages require models with different structures. A sample of the modeling r^2values is given in Table 3, which shows results for utterances of the form *miC(C)a*.[4]

The models are named for the speaker(s) they were developed for; r^2 values for speakers' own models are in boldface in the table. Variables were the phasing relations between pairs of gestures; the stiffness of the vowel gestures was also varied between utterances with single and geminate consonants for models #J3, I1 and I2&3.

This kind of modeling has the advantage of capturing the numerous measured differences between utterances with single and geminate consonants with relatively few parameterized differences, reducing much variability to a few interpretable differences. The similarity of the models obtained for the three Italian speakers suggests that they were all showing similar temporal organization, despite the superficial differences among them.

Table 3. *Fits of the durational differences between utterances with single and geminate consonants for the intervals between articulatory events. The entries in the table are the r^2 values between the differences between the measured interval durations in single and geminate utterances and the differences predicted by the models developed for each of the speakers. Results for utterances of the form miC(C)a are shown here.*

	Japanese-style models (V-and-C)			Italian-style models (V-V)	
speaker	J1	J2	J3	I1	I2&3
J1	**.91**	.57	.89	.31	.26
J2	.69	**.92**	.68	.74	.73
J3	.90	.53	**.99**	.02	.01
I1	.02	.00	.03	**.93**	.80
I2	.15	.00	.23	.66	**.88**
I3	.20	.24	.16	.83	**.96**

4.1 Gestural representation of geminates

Since the two structures for temporal organization are being contrasted with respect to differences due to consonantal length contrast, the criterion of goodness-of-fit for the models was how accurately they predicted the pattern of differences between single and geminate utterances. Thus a crucial aspect of the models is how the single/geminate contrast is represented, in two ways—the specific consonantal gesture(s) and the differences in the utterance as a whole that result from this contrast. The modeling procedure outlined above assesses the overall validity of the representation; it remains to consider how the length contrast should be modeled in the consonantal gesture.

In both the models shown in Figure 7, a geminate consonant was modeled as one gesture with different parameter values from those for the gesture for a single consonant. In both languages the interval during which the lips or tongue tip remained in a closed position was significantly longer for the geminate than for the single consonant. This pattern was modeled by having the gesture for a geminate consonant remain active for a longer time after reaching its target position than the gesture for a single consonant.

For the Japanese speakers, and Italian speakers I1 and I3, in addition, the lips moved more slowly when forming the constriction for a geminate than a singleton, so the duration of the movement towards the target was significantly longer. This difference was modeled by reducing the stiffness of the gesture for the geminate versus the single consonant. Both the longer period of activation and the decreased stiffness have the effect of increasing the duration of the gesture for a geminate. These parameter changes are schematized in Figure 8. These models require specification of the characteristics of the movement forming a constriction and, for consonants, the duration of the period of activation. For vowels, the end of the period of activation was not specified; how long the gesture remains active depends on the timing of the following gestures.

The models presented here, which represent geminate consonants by a single gesture, offer an economical way of getting a good correspondence with the articulatory data. While the goal of the modeling was to achieve a good fit of the data, ideally the models should reflect the phonological structure of the utterances being modeled. Conceptually, a model using two gestures to represent a geminate may be more appealing. In Articulatory Phonology, phonological contrasts (such as presence or absence of voicing) are represented by the presence or absence of a gesture. Thus, adding a gesture to geminate a consonant would coincide better with the representation of other kinds of phonological contrasts. This seems appropriate for languages such as Japanese and Italian, in which consonant length is phonologically contrastive. A two-gesture representation also reflects the fact that phonologically, geminates often seem to behave like two units (see e.g., Schein & Steriade, 1986).

In a two-gesture representation of geminates, the parameters (stiffness and phasing) for each of the two gestures in a geminate would have to be specified, as for any other gesture. If each of these gestures had the same parameter values as the gesture for a single consonant, then the total duration of the geminate would be greater than that of a single consonant. But exactly how much greater would depend on the phasing between the two gestures. This phasing relation might well vary between languages. Figure 9 illustrates two possible phasing relations between the two gestures of a geminate. The total duration of the geminate would, of course, be greater in the lower of the two diagrams in Figure 9. However, in order for two gestures to form a geminate, rather than two separate consonants, they must be timed in such a way that the articulators maintain the target configuration of the vocal tract.

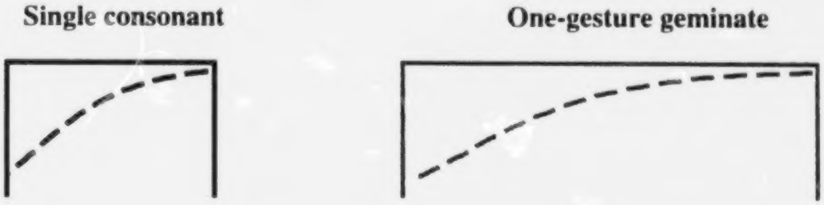


Figure 8. Hypothetical gestural representation (box) and resulting trajectory of a tract variable towards its target (dashed line) for gestures associated with single and geminate consonants, with the geminate consonant modeled as a single gesture having lower stiffness and a longer period of activation than the single consonant.

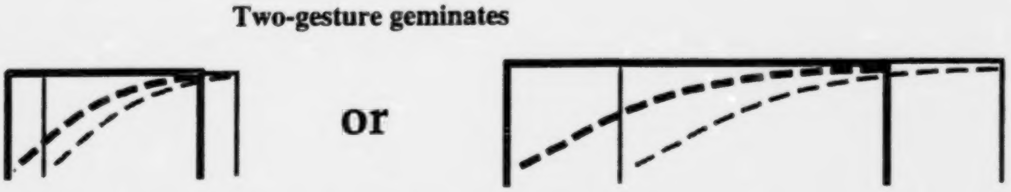


Figure 9. Model of a geminate consonant as two gestures. The figure illustrates a hypothetical gestural representation (box) and resulting trajectory of a tract variable towards its target (dashed line) for two alternate phasing relations between two gestures of the geminate.

In the right-hand diagram of Figure 9, the delay between the two gestures may be too great to maintain the target configuration. The extent of overlap between the two hypothetical gestures cannot be determined from the data: in the movement traces the geminates, like the singletons, had a single articulatory maximum, i.e. they showed up as a single hump. Thus relating this two-gesture model to the articulatory data is problematic.

The most constrained form of the two-gesture model would be to assume that each of the two gestures has the same parameters as the gesture for a single consonant. However, recall that for most of the speakers the movement to form a geminate closure was significantly slower than for a single closure. This implies that if a geminate consists of a pair of gestures, at least the first one would have to have a lower stiffness than if it were alone. Thus, even for a two-gesture model, the dynamic parameters for a geminate consonant have to be different from the parameters for a single consonant. This means that a two-gesture model effectively requires two changes to represent a single phonological contrast, rather than the one change (the parameter values) that is required by the one-gesture representation. For this reason, the two-gesture model appears to be a more costly approach.

4.2 Generalization of results for geminates to utterances with intervocalic clusters

Because the single and geminate consonants contrast only in the time domain, they were used in this study for the comparison of the timing structures of Japanese and Italian. However, to ensure that the observed differences between the utterances with single and geminate consonants were general effects of durational differences, and not particular to geminates, utterances with the homorganic cluster /mp/ were also measured and compared to the utterances with geminates.[5] There were very few statistically significant differences in the durations of the various measured intervals, and where there were differences, they were similar in magnitude to the small differences that had been observed between oral and nasal geminates—that is between /pp/ and /mm/, /tt/ and /nn/. In general, the durations of the measured intervals in utterances with the intervocalic cluster seemed to be mostly dependent on the nasality of the adjacent part of the cluster: that is, measures relating to the first part of the utterance tended not to differ between /mp/ and /mm/, and measures of the second part tended to pattern similarly in /mp/ and /pp/. This suggests that the cluster does not differ from the geminates in any way relating to length, but that it does constitute a sequence with respect to nasality.

The /mp/ cluster tested in this experiment could be represented as a single labial gesture with co-ordinated velic opening and closing gestures. However, heterorganic clusters would have to be specified using two or more oral gestures; thus in general, clusters cannot be distinguished from single consonants merely by altering the parameters of a single gesture. Therefore, if a single-gesture representation is adopted for geminates, it could not be extended to clusters. This restriction seems undesirable, since it appears that the cluster and the geminates pattern in the same way with respect to durational effects. This similar patterning of clusters and geminates supports the proposal for representing geminates as two gestures, rather than one.

5 CONCLUSION

The results presented here show how minimal manipulation of structural relations organizing dynamic primitives (gestures) can give rise to complex, inter-related surface timing patterns. One of the principal advantages of Articulatory Phonology is that the intrinsic duration of gestures facilitates the representation of temporal relations. The patterns of temporal organization observed among articulatory gestures can vary between languages, but seem to vary in a way that corresponds to the traditional descriptions of

languages' rhythms, and can be described in terms of how different gestures are coordinated in time. Steriade's (1990) work has also suggested that some phonological processes may be interpretable as changes in the phasing of consonants and vowels relative to each other. In the framework of Articulatory Phonology, the account of the patterning of such phonological processes is related to cross-linguistic differences in rhythmic units and durational patterns. Different structural relations among gestures are one of the ways that languages create different rhythms.

REFERENCES

Abbs, J. H., & Nadler, R. D., (1987). *User's Manual for the University of Wisconsin X-Ray Microbeam.* Madison, WI: Waisman Center.

Archangeli, D. (1985). Yokuts harmony: coplanar representation in nonlinear phonology. *Linguistic Inquiry, 16,* 335-372.

Bertinetto, P. M. (1983). Ancora sull'italiano come lingua ad isocronia sillabica. *Scritti Linguistici in onore di Giovan Battista Pellegrini II* (pp. 1073-1082). Pisa: Pacini.

Browman, C. P., & Goldstein, L. (1986). Towards an articulatory phonology. *Phonology, 3,* 219-252.

Browman, C. P., & Goldstein, L. (1990a). Tiers in articulatory phonology, with some implications for casual speech. In J. Kingston & M. E. Beckman (Eds.), *Papers in Laboratory Phonology I: Between the grammar and the physics of speech.* (pp. 341-376). Cambridge, MA: Cambridge University Press.

Browman, C. P., & Goldstein, L. (1990b). Gestural specification using dynamically-defined articulatory structures. *Journal of Phonetics, 18,* 299-320.

Browman, C. P., & Goldstein, L. (1992). Articulatory phonology: An overview. *Phonetica, 49,* 155-180.

Dauer, R. M. (1983). Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics, 11,* 51-62.

Fowler, C. A. (1980). Coarticulation and theories of extrinsic timing control. *Journal of Phonetics, 8,* 113-133.

Fowler, C. A. (1981). A relationship between coarticulation and compensatory shortening. *Phonetica, 38,* 35-50.

Fowler, C. A. (1983). Converging sources of evidence on spoken and perceived rhythms of speech: Cyclic production of vowels in monosyllabic stress feet. *Journal of Experimental Psychology: General, 112,* 386-412.

Goldstein, L. (1989). On the domain of the Quantal Theory. *Journal of Phonetics, 17,* 91-97.

Joos, M. (1948). *Acoustic Phonetics* (Language Monographs No. 23). Baltimore: Linguistic Society of America at the Waverly Press.

McCarthy, J. J. (1981). A prosodic theory of nonconcatenative morphology. *Linguistic Inquiry , 12,* 373-418.

McCarthy, J. J. (1982). Prosodic templates, morphemic templates, and morphemic tiers. In H. van der Hulst & N. Smith (Eds.) *The structure of phonological representations, Part I* (pp. 191-224). Dordrecht: Foris.

McCarthy, J. J. (1989). Linear ordering in phonological representation. *Linguistic Inquiry, 20,* 71-99.

Nadler, R. D., Abbs, J. H., & Fujimura, O. (1987). Speech movement research using the new x-ray microbeam system. *Proceedings of the Eleventh International Congress of Phonetic Sciences* 1 (pp. 221-224). Tallinn: Academy of Sciences of the Estonian SSR.

Öhman, S. E. G. (1966). Coarticulation in VCV utterances: spectrographic measurements. *Journal of the Acoustical Society of America, 39,* 51-168.

Öhman, S. E. G. (1967). Numerical model of coarticulation. *Journal of the Acoustical Society of America, 41,* 310-320.

Prince, A. (1987). Planes and copying. *Linguistic Inquiry, 18,* 491-509.

Saltzman, E. (1986). Task dynamic coordination of the speech articulators: a preliminary model. *Generation and modulation of action patterns* (pp. 129-144). Experimental Brain Research Series 15. New York: Springer-Verlag.

Saltzman, E., & Kelso, J. A. S. (1987). Skilled actions: a task-dynamical approach. *Psychological Review, 94,* 84-106.

Saltzman, E., & Munhall, K. G. (1989). A dynamical approach to gestural patterning in speech production. *Ecological Psychology, 1,* 333-382.

Schein, B., & Steriade, D. (1986). On geminates. *Linguistic Inquiry, 17,* 691-744.

Smith, C. (1992). *The temporal organization of vowels and consonants.* Unpublished doctoral dissertation, Yale University.

Steriade, D. (1986). Yokuts and the vowel plane. *Linguistic Inquiry, 17,* 29-146.

Steriade, D. (1990). Gestures and autosegments: comments on Browman and Goldstein's "Gestures in Articulatory Phonology." In J. Kingston & M. E. Beckman (Eds.), *Papers in Laboratory Phonology I: Between the grammar and the physics of speech* (pp. 382-397). Cambridge: Cambridge University Press.

Vance, T. J. (1987). *An introduction to Japanese phonology.* Albany, NY: State University of New York Press.

Westbury, J. R. (1991). The significance and measurement of head position during speech production experiments using the X-ray microbeam system. *Journal of the Acoustical Society of America, 89,* 1782-1791.

FOOTNOTES

[*] *Papers in Laboratory Phonology 4,* in press.

[†] UCLA Division of Head & Neck Surgery

[1] Only two tongue pellets could be used with speaker I3, so the horizontal and vertical movements of the rearmost pellet were measured.

[2] Each movement trace was analyzed individually, and separate analyses of variance were performed for each vowel pattern (a-i and i-a) for each speaker. The factors in these analyses were length and nasality of the intervocalic consonant. The significance level for results reported here was $p < .05$.

[3] Non-significant in the vertical movement of the Tongue Body for speaker I3.

[4] Exceptionally, Table 3 shows that the Italian models fit this set of utterances from Japanese speaker J2 quite well: this goodness-of-fit was not found in the *maC(C)i* utterances.

[5] It was not possible to make a direct statistical comparison between the utterances with single consonants and those with clusters.

BEST COPY AVAILABLE

Haskins Laboratories Status Report on Speech Research
1993, SR-115/116, 57-68

Divergent Developmental Patterns for Infants' Perception of Two Non-Native Consonant Contrasts*

Catherine T. Best,† Gerald W. McRoberts,‡ Rosemarie LaFleur,††† and Jean Silver-Isenstadt††††

Young infants discriminate non-native and native consonant contrasts, yet 10-12 month olds discriminate most non-native contrasts poorly, like adults. However, English-speaking adults and 6-14 month infants discriminate Zulu clicks, consistent with a model predicting that listeners who have a native phonology assimilate non-native consonants to native categories when possible, but hear Non-Assimilable (NA) consonants as nonspeech sounds (Best, McRoberts & Sithole, 1988). NA contrasts thus avoid language-specific effects and are discriminated, whereas consonants assimilated equally into a Single Category (SC) are discriminated poorly by listeners showing language-specific influences; other possible assimilation patterns show poor to excellent discrimination. This study directly compared discrimination of NA clicks and SC ejectives by 6-8 and 10-12 month olds with a conditioned fixation habituation procedure. Consistent with predictions, the younger group discriminated both non-native contrasts and a control English contrast, whereas the older group discriminated only the NA and English contrasts.

INTRODUCTION

The influence of language experience on speech perception is evident in the limitations that have been observed in adults' categorization and discrimination performance with phonetic distinctions that do not contrast phonologically in their own language(s) (e.g., Best & Strange, 1992; Flege, 1989; Flege & Eefting, 1987; Lisker & Abramson, 1970; MacKain, Best, & Strange, 1981; Miyawaki et al., 1975; Polka, 1991, 1992; Tees & Werker, 1984; Trehub, 1976; Werker & Logan, 1985; Werker & Tees, 1984a). Yet given that infants learn whichever language is used in their homes within their first few years, they obviously must be able, from fairly early on, to perceive virtually the full range of phonetic contrasts used in any of the world's languages. Research with infants under about 6 months has borne out this near-universal phonetic sensitivity for consonant and vowel contrasts. Such young infants discriminate segmental contrasts regardless of whether they occur in the native language or only in unfamiliar languages (e.g., Eilers & Minifie, 1975; Eilers, Wilson, & Moore, 1977; Eimas, 1975; Eimas & Miller, 1980; Eimas, Siqueland, Jusczyk & Vigorito, 1971; Jusczyk & Thompson, 1978; Lasky, Syrdal-Lasky, & Klein, 1975; Streeter, 1976; Swoboda, Morse & Leavitt, 1976; Swoboda, Kaas, Morse & Leavitt, 1978; Trehub, 1973, 1976). This striking developmental difference in perception of non-native phonetic contrasts indicates that sometime between early infancy and adulthood the listener's experience with a particular language comes to exert a powerful influence on speech perception.

We would like to thank the Wesleyan students who served as research assistants during the course of conducting this study: Shama Chaiken, Laura Klatt, Pam Spiegel, Amy Wolf, David Fleishman, Leslie Turner, Ashley Prince, Ritaelena Mangano, Pia Marinangelli, Heidi Queen, Peter Kim, and Jane Womer. We also thank Janet Werker for the loan of her Nthlakampx (Thompson Salish) and English stimulus materials, and for numerous helpful discussions with her about the findings. We are especially grateful to the parents of our infant subjects for their willingness to let their infants participate, and for their interest in the research.

This work was supported by NIH grants HD01994 to Haskins Laboratories and DC00403 to the first author. The research was begun when the second author was a doctoral student at the University of Connecticut and a research assistant at Haskins Laboratories.

An important theoretical issue for developmental cross-language research to resolve is the nature of this language-specific effect: When and why does the ambient language begin to leave its mark on speech perception, particularly on the perception of non-native sound patterns? Regarding the first part of the question, a number of studies indicate that language-specific perceptual effects appear before the end of the infant's first year. A possible clue to the second part is that the timing of these early perceptual changes varies for different aspects of sound patterning in speech (for in-depth discussions of possible developmental accounts, see Best, in press a; Jusczyk, 1993, in press; Werker & Pegg, 1992). To summarize, Werker and her colleagues have provided strong evidence that a native-language effect on perception of consonant contrasts becomes established between 8 and 10 months of age. After 10 months, English-learning infants no longer discriminate several non-native consonant contrasts from the Hindi and Nthlakampx (a Salish language [Thompson] of the Canadian Pacific region) languages which they can clearly discriminate prior to 8 months (Werker, Gilbert, Humphrey, & Tees, 1981; Werker & Tees, 1984b; Werker & Lalonde, 1988). Certain language-specific effects appear even earlier for vowels. English and Swedish 6 month olds each show internally-organized perceptual categories only for the vowels categories of their own language, i.e., poor discrimination of "good" tokens in the neighborhood of the category prototype but relatively better discrimination among "poor" tokens in the category periphery (Kuhl et al., 1992). And although English-learning 4-1/2 month olds can discriminate two non-English vowel contrasts from German , 6-8 month olds show a "magnet effect" asymmetry in discriminating the less vs. more English-like vowels in these German contrasts, and 10-12 month olds fail to discriminate them altogether, in either direction (Polka & Werker, 1994). Infants' attunement to some more global prosodic properties of speech may be evident even earlier than that for vowels (Mehler et al., 1988); nonetheless their attunement to certain other prosodic properties may not appear until their second half-year (Jusczyk et al., 1992). Regardless of onset age differences for these diverse aspects of speech, the findings clearly suggest that sensitivity to specific phonetic properties in speech declines if the language environment does not provide exposure to them.

This pattern of results led some to propose that exposure to specific phonetic contrasts during an early critical period is needed to maintain the neural elements that are innately tuned to the phonetic features involved, and conversely, that lack of exposure to particular contrasts results in attrition of the associated neural elements (e.g., Aslin & Pisoni, 1980; Eimas, 1975). Alternatively, it has been suggested that differential phonetic experience may sharpen attention or psychoacoustic responsiveness to phonetic properties found in the native language and/or may attenuate such responsiveness to properties that are absent from that language (e.g., Burnham, 1986; Diehl & Kluender, 1989; Lively, Logan, & Pisoni, 1993; Logan, Lively, & Pisoni, 1991; Pisoni, Aslin, Perey, & Hennessy, 1982; Walley, Pisoni, & Aslin, 1981). Still others have argued that, instead, differential phonetic experience shapes the higher-level processing (e.g., phonological coding, retention in memory) of auditory information from the speech signal (e.g., Tees & Werker, 1984; Werker & Tees, 1984a).

As has been pointed out elsewhere, a simple sensorineural loss explanation is untenable for several reasons (e.g., Best, 1984, 1994; MacKain, 1982; Werker & Tees, 1984a). For one, adults' perception of non-native phonetic contrasts can at least sometimes be improved by learning the other language (e.g., Flege, in press; MacKain, Best & Strange, 1981; Williams, 1979) or by laboratory training (e.g., Lively, Logan, & Pisoni, 1993; Logan, Lively, & Pisoni, 1991; Pisoni et al., 1982). In addition, discrimination of non-native contrasts benefits from task manipulations that reduce memory demands (e.g., Carney, Widin, & Viemeister, 1977; Werker & Logan, 1985), or that isolate the crucial acoustic cues by removing them from speech context (e.g., Miyawaki et al., 1975; Werker & Tees 1984a). Moreover, in some cases listeners have had exposure to the phonetic properties of non-native contrasts on which they have shown perceptual difficulties, because those phonetic features occur in allophonic variants of native phonological categories (e.g., MacKain, 1982). None of these observations is consistent with sensorineural attrition due to lack of exposure during an early critical period.

Of the remaining two accounts, the attentional one may be weakened, although not refuted, by reports that training or instructional manipulations which focus listeners' attention on the critical acoustic properties of non-native contrasts fail to improve perceptual performance on the associated phonetic contrasts within speech contexts (e.g., Werker & Tees, 1984a). Such findings suggest that the higher-level processing explanation accounts for language-specific effects on speech perception better than does the attentional explanation. Alternatively, however, the failures may simply indicate that the attentional manipulations were inadequate for the attentional focus on the isolated cues to carry over to acoustically complex syllabic contexts.

Note that both the attentional account and the information processing account nonetheless seem to assume, explicitly or implicitly, that lack of experience with specific phonetic features or contrasts lies at the root of the ubiquitous difficulty adults have with non-native phonological contrasts. That is, they presume that phonetic experience *per se* is the source of the language-specific perceptual effects that emerge in infancy. This assumption was called into question by the recent finding that monolingual English-speaking adults, and infants up to the oldest age tested (14 months), showed good discrimination of click consonant contrasts from Zulu (Best, McRoberts, & Sithole, 1988). Clicks are produced by making the tongue form a complete closure around the palate (roof of the mouth), then causing a small vacuum to form by drawing the side or tip of the tongue downward. Ultimately, the tongue breaks its contact with the palate at that point and the vacuum is released, producing a suction sound or click. Click sounds fall entirely outside the range of allophonic experience with spoken English Yet even without training and without any lowering of task demands, adults performed much better on the clicks than they had been reported to do on other non-native contrasts, whether or not the phonetic properties of those other contrasts coincide to any extent with native allophonic experience (see, e.g., Tees & Werker, 1984). Moreover, there was no developmental decline in infants' discrimination of the clicks, contrary to the marked decline at 10-12 months for discrimination of the non-native contrasts tested by Werker and colleagues.

The findings with click consonants suggest that the effect of experience on perception of non-native contrasts is not a simple effect of experience, or lack thereof, with specific phonetic features or contrasts in speech. Rather, language-specific perceptual effects must reflect listeners' knowledge of the *relation* between physical, phonetic properties in speech and the more abstract linguistic functions that phonological categories and contrasts serve in the native language. The phonetic properties of the other non-native contrasts tested, but not of the clicks, apparently made them susceptible to being perceived in some relation to native phonological categories. Best and colleagues (1988) posited that listeners who have become familiar with the phonological system of a specific language tend to perceptually assimilate unfamiliar non-native consonants and vowels to their own phonological categories based on phonetic similarities, *if* the similarities are sufficient to permit this. On the other hand, if particular non-native sound patterns deviate too greatly from the phonetic properties employed in the native phonological system—e.g., the suction-release action for the click consonants is unlike any of the phonetic features that comprise the English phonological system—listeners should fail to assimilate those sounds as potential phonological elements. In the latter event, they will instead perceive the non-assimilated speech elements as nonspeech sounds, as did the English-speaking adults who heard the Zulu clicks (Best et al., 1988).

Those suppositions form the basis of a Perceptual Assimilation Model (PAM), which is more fully described elsewhere (Best, 1993, 1994, in press a, b; Best & Strange, 1992). Its primary contribution to the empirical literature to date is that it accounts at once for both the high level of discrimination for non-native click contrasts, on the one hand, and the more commonly-reported adult perceptual difficulties and developmental decline in infant discrimination for other non-native consonant contrasts, on the other hand. The model has broader theoretical implications, however. PAM makes systematic predictions about other types of non-native contrasts, *viz* that discrimination levels should range from poor to excellent, depending on differences in the way the phonetic properties of non-native phonetic segments (consonants and vowels) are assimilated to native phonological categories.

To summarize, the phonetic properties of a non-native segment may bias it toward perceptual assimilation into the phonological system of the listener's native language. If assimilated into a particular native category, it may either match the ideal phonetic representation of the category, it may deviate modestly from that ideal but be heard as a good exemplar of the category, or it may fall near the category's periphery and be heard as a relatively poor exemplar of the category. Alternatively, the phonetic properties of a non-native segment may fall somewhere in between native categories, in "uncommitted phonetic space," such that it is heard as a speech sound (i.e., potential phonological element) but it is not assimilated into any specific native category. Finally, the phonetic properties of a non-native segment may be so uncharacteristic of those employed in the native phonological system that it is not assimilated as a speech sound, but instead falls outside the phonological realm altogether and is perceived to be a nonspeech sound (i.e., environmental sound or nonlinguistic human sound such as a cough, hiss, or a disapproving "tsk-tsk"). To English speakers, the click consonants of Zulu fail to be assimilated as potential elements of a phonological system, and are heard as nonspeech (Best et al., 1988).

Predicted discrimination levels for non-native contrasts follow from the assimilation patterns of each of the contrasting segments. When two non-native segments are assimilated into a single native category, discrimination should be poor if both fall equally close to the native category ideal, a case referred to as Single Category assimilation (abbreviated **SC**). It has been argued that many of the non-native contrasts for which adults and older infants have been reported to show perceptual difficulties are likely to be SC contrasts (e.g., Best, 1993, 1994, in press a, b). On the other hand, when two non-native segments are assimilated into a single native category, but unequally such that one is close to the native ideal while the other is in the category periphery, listeners should perceive a Category Goodness difference (**CG**) between them. In CG contrasts, discrimination is relatively good, the exact level depending on the magnitude of the goodness difference between the two sounds and their proximity to the periphery of the native category. Discrimination should be even better, approaching native listener levels, when the contrasting non-native segments are assimilated to Two Categories (TC) in the native phonology. One or both of the non-native segments may instead fall in Uncommitted phonetic space (UC or UU, respectively), leading to relatively good discrimination in the first case or moderate to poor discrimination in the second. Finally, the contrasting non-native categories may be Non-Assimilable (NA) with respect to the native phonological system, as described above for the Zulu clicks, in which case they should be discriminated moderately to very well (for more detailed description, see Best, in press a, b).

A small number of studies has examined PAM's predictions for adult perception of non-native contrasts (Best et al., 1988; Best & Strange, 1992; Polka, 1991, 1992, submitted). Their findings have supported the model's predictions (all types of non-native contrasts have been tested, except those with assimilation to uncommitted phonetic space). However, extending PAM to explain language-specific developmental changes in infant speech perception is problematic at present. There has been only one published report on infants, which looked only at the NA assimilation type. Most importantly, the NA type has not been compared to an assimilation pattern for which a developmental decline in discrimination would instead have been predicted (Best et al., 1988). Moreover, comparison of the click findings to other cross-language infant studies are confounded by a difference in methodology. Whereas Werker's studies employed the conditioned head-turn response in the multi-trial go/no go procedure used at a number of infant speech perception laboratories (e.g., Eilers, Wilson, & Moore, 1977; Kuhl, 1980), Best and colleagues (1988) used a conditioned visual fixation response in an infant-controlled habituation-dishabituation procedure (Miller, 1983). The conditioned fixation procedure had not been used previously in tests of infant consonant perception. It is at least plausible that it may be cognitively less demanding and/or psychophysically more sensitive than the conditioned head-turn procedure. If so, the divergence between the 10-12 month decline in discrimination of Werker's Hindi and Nthlakampx stimuli and the lack of developmental change in discrimination of Zulu clicks might be attributable solely to the difference in methodology.

Therefore, it was important to verify the robustness of the NA pattern for click contrasts, and test the contrary prediction of developmental decline for SC contrasts, in the same infants and using a single methodology. For this purpose, we used the conditioned fixation procedure to test 6-8 month old and 10-12 month old English-learning American infants on three contrasts: native English /ba/-/da/, Nthlakampx velar vs. uvular ejectives /k'æ/-/q'æ/, and Zulu voiceless unaspirated apical vs. lateral clicks /ǀa/-/ǁa/. These ages were tested because previous reports indicated that the younger age should discriminate native and non-native consonant contrasts without difficulty, whereas the older age should show marked difficulty in discriminating non-native consonant contrasts other than the clicks. The English and Nthlakampx stimuli were those used by Janet Werker (Werker & Tees, 1984a, b). The Zulu click stimuli were those used Best and colleagues (1988). The English contrast served as a native control comparison. The clicks had met the criteria for an NA assimilation type according to the adult findings of Best et al., and had shown good discrimination across all ages, without a developmental decline in infants' discrimination. The Nthlakampx ejectives were expected to fit the pattern for SC assimilation, whereby adult English listeners tend to assimilate both the velar and the uvular ejective as equally "odd" exemplars of the English voiceless stop /k/. The SC assimilation prediction for infants was that there should be good discrimination at 6-8 months but poor discrimination at 10-12 months. Such a pattern would replicate the Werker and Tees (1984b) findings in a different laboratory and with a different infant testing procedure.

The Nthlakampx contrast in particular was selected for several reasons. The perceptual results for both adults and older infants differ dramatically between this poorly-discriminated ejective contrast and the easily-discriminated click contrast. Nonetheless, these ejectives and the voiceless unaspirated clicks show a number of similar acoustic properties (see Best et al., 1988; Werker & Tees, 1984a). Both types of consonants produce brief noise bursts that are higher in amplitude than the following vowel, and show similar high frequency poles in the noise spectrum around 4200-4600 Hz. The noise bursts for both consonant contrasts are separated from the subsequent vowel by a brief silent interval, thus the vowel is unlikely to produce masking of the noise in either case (see Werker & Tees, 1984a). There are, of course, some acoustic differences between the ejectives and the clicks. The click noise bursts are 9 m longer ($M = 47.4$ ms) than the bursts for the ejectives ($M = 38.5$ ms). However, the total duration of click + silence ($M = 66.9$ msec) is equal to that for ejective burst + silence because the silent interval is shorter for clicks ($M = 19$ msec) than for ejective bursts ($M = 29$ ms). Thus there is slightly more likelihood of vowel masking for the clicks, which would work *against* good discrimination. The noise bursts differ between the two clicks much more strikingly in the frequency of the lower spectral pole (120 Hz vs. 2450 Hz) than do those of the ejectives (3100 Hz vs. 3200 Hz). The noise bursts for the two types of consonant contrasts most likely also differ strongly in their amplitude envelopes.

Method

Subjects. Twenty-four infants were included in the study, 12 at 6-8 months (9 males, 3 females; M_{age} = 6 mo. 18 days, range = 5 mo. 30 days to 7 mo. 26 days) and 12 at 10-12 months (5 males, 7 females; M_{age} = 11 mo. 7.5 days, range = 9 mo. 26 days to 12 mo. 14 days). All were normal, full-term infants without gestational or labor/delivery complications, and were free of ear infections or colds on the day of testing. An additional 39 infants were excluded from the final data set (crying = 16; equipment problems = 6; experimenter error = 3; inattentiveness = 13 [i.e., 10 or more consecutive trials without fixation responses]; Down syndrome = 1).

Stimulus materials. The three stimulus contrasts used in this study were the English /ba/-/da/, Nthlakampx ejectives /k'æ/-q'æ/, and Zulu clicks /ǀa/-/ǁa/. All stimulus contrasts included multiple natural tokens produced by a native speaker of the language involved, selected for similarity in duration, amplitude and frequency characteristics of the tokens within the pairs of contrasting categories. The English and Nthlakampx contrasts were produced by male adult speakers; the Zulu contrast was produced by an adult female. Acoustic measurements of the non-native contrasts are reported in the original papers (Best et al., 1988; Werker & Tees, 1984a, b).

Procedure. We employed the same conditioned visual fixation habituation procedure used in our previous study (Best et al., 1988; see also Miller, 1983). In this procedure, tokens of one speech category were played to the infant over a hidden loudspeaker at a conversational listening level whenever the infant fixated on a rear-projected picture (colored checkerboard) presented on screening material affixed to a window in the wall they face during testing. A video monitor connected to a hidden camera at the side of the projection window displayed the infant's head and shoulders to an observer in the adjacent room, who registered the infant's fixations (as well as bouts of crying and sleeping) via key press input to a computer (Atari 800). The computer registered the fixation duration, and controlled the presentation of audio stimuli from a reel-to-reel tape deck (Otari 5050 MXB). When the infant looked away from the picture, the observer released the "looking" key and the computer stopped the presentation of the speech sounds to the infant.

Trial duration was under infant control: if the infant looked away from the slide for two consecutive seconds the trial ended and the slide blankened during the intertrial interval (ITI). After one second the slide automatically reappeared, beginning the next trial. Habituation was defined as two consecutive trials with fixation durations below 50% of the mean of the two highest preceding trials (Miller, 1983). The criterion was calculated and updated on a trial-to-trial basis by the computer program. Once habituation was met during the first phase, referred to as the familiarization phase, audio presentations shifted to the contrasting speech category for the test phase, which continued until the infant again habituated. The index of discrimination is any change in fixation during the first two test trials relative to the last two familiarization trials. Full technical details for the procedure are available in the original report (Best et al., 1988).

During testing the infant sat in an infant seat or on the parent's lap, about 3 feet from the rear-projection window, in the dimly lit testing room. Both were seated in a small booth constructed by attaching two partitions to the wall on either side of the projection window, about 3.5 ft. apart. Each partition was approximately 6 ft. high, and extended 4 ft out from the wall. The booth was open at the back, and its sides were covered with black fabric. The wall at the front of the booth was also covered with black fabric, except for the 2 ft. × 2 ft. area directly in front of the infant's head where the picture was projected. A Jamo loudspeaker was used for stimulus presentations; it was attached to the wall 3 ft. above the projection window, and was hidden behind the black cloth covering the wall. Speech was presented to the infant at a 65-70 dB sound pressure level. Both the parent and the infant observer listened to music over closed-design headphones (Sennheiser HD440) during testing to prevent them from inadvertently biasing the infant's behavior or the fixation observations.

Each infant completed all three speech discrimination tests within a single session. Test order was randomized across infants within each age group. Short breaks of 5-10 minutes were taken between tests if necessary to maintain infants' attention and/or to soothe them if they had become irritable. Otherwise, the session proceeded from one test to the next with only the 1-2 minute break needed for re-positioning the audio tape and restarting the computer program. Infants were eliminated from the final data set if they cried for more than a cumulative 30 seconds during any test, or if they cried during any of the trials just before or after the test shift.

Results

Inter-observer reliability. The data for a random selection of seven of the infant subjects (i.e., 21 individual tests) were rescored by second observers re-running the testing program while viewing the videotapes. Thus, interobserver reliability was assessed off-line. Reliability was quite good ($r = .91$ to $.985$).

Habituation during Familiarization. Habituation during the familiarization phase of the tests was verified by analyses of variance (ANOVAs) on both forward and backward habituation curves. Forward habituation analyses compared the mean fixation in the first two trials against the mean in the final two trials prior to the stimulus shift in all tests, in an Age (2) × Language (3) × Trial Block (2) ANOVA. The Trial Block effect revealed a significant decline in fixation from the first familiarization trials ($M = 12.36$ s) to the final trials before the test shift ($M = 2.46$ s), $F(1, 22) = 87.476$, $p < .0001$. The Age effect was also significant, $F(1, 22) = 6.081$, $p < .025$, indicating that the younger group looked significantly longer during familiarization ($M = 9.02$ s) than did the older group ($M = 5.8$ s). However, an Age × Trial Block interaction showed this age difference to be restricted to the beginning trials of the familiarization phase (Ms = 15.48 and 9.25 s, respectively); both ages habituated to the same low fixation level by the final preshift trials (Ms = 2.56 and 2.36 s, respectively), $F(1, 22) = 8.104$, $p < .01$.

A separate Age (2) × Language (3) × Trials (4) ANOVA was conducted on the backward habituation data, for the last four trials of familiarization. The Trials effect showed a dramatic and significant decline in fixations during the last two preshift trials (trials -2 and -1: Ms = 2.67 and 2.25 s) relative to the two trials just preceding those (trials -4 and -3: Ms = 11.27 and 13.45 s), $F(3, 66) = 18.313$, $p < .0001$ (see Figure 1). A significant Age effect, $F(1, 22) = 5.68$, $p < .03$, indicated that the younger infants fixated longer during these familiarization trials ($M = 9.06$ s) than did the older infants ($M = 5.76$ s). This age difference was evident only during the -4 and -3 trials before the shift to the test phase (6-8 month Ms = 13.14 and 18.0 s; 10-12 month Ms = 9.40 and 8.91 s). As the forward habituation analysis had shown, fixation had dissipated to the same low fixation levels for both ages by the two trials just preceding the shift (6-8 month Ms = 2.97 and 2.20 s; 10-12 month Ms = 2.42 and 2.30 s).

Given that both forward and backward habituation values are constrained by the habituation criteria we used, we also examined possible language and age differences in the number of trials to habituation, and in mean looking time per habituation trial in separate Age (2) × Language (3) ANOVAs. These two indices are not constrained by the method we used for determining habituation. No main effects or interactions approached significance in either of the latter ANOVAs.

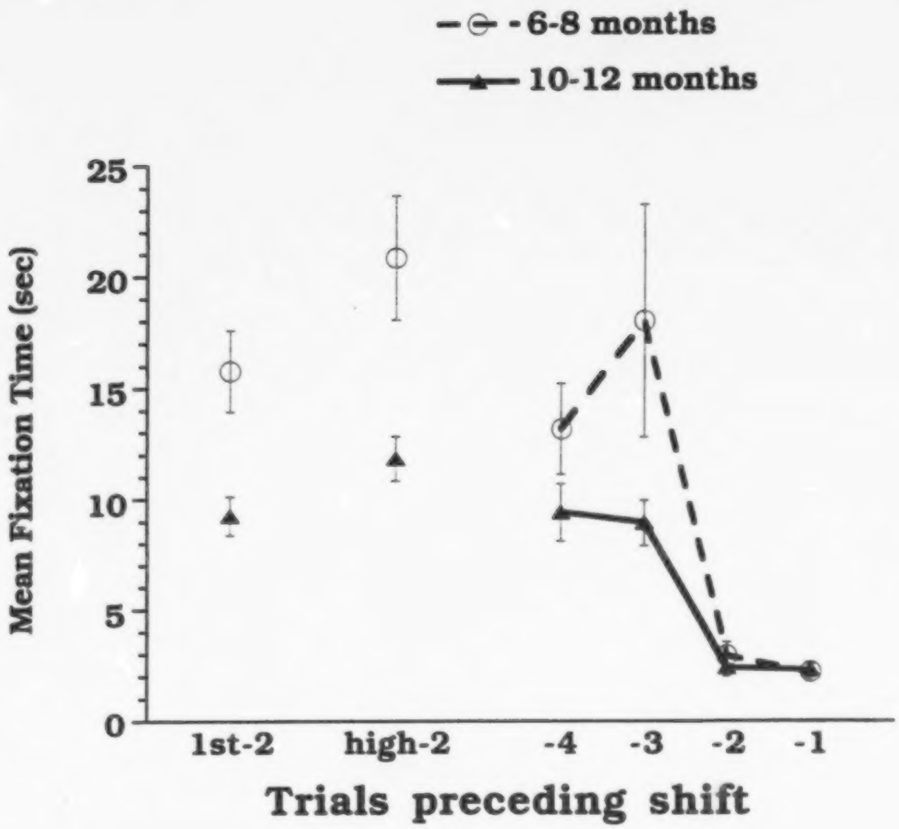


Figure 1. **Backward habituation curves for the last 4 trials of the familiarization phase, with standard error bars, shown separately for 6-8 and 10-12 month olds. For comparison, the mean fixation time for the first two trials and the mean for the two highest trials are also displayed.**

Discrimination results. Discrimination was assessed by comparing the mean fixation duration during the last two trials of the familiarization phase (Preshift Trial Block) against the mean fixation during the first two trials of the test phase (Postshift Trial Block). The postshift block was defined as beginning with the first trial after the stimulus shift in which the infant fixated on the slide and thus had an opportunity to begin hearing the test stimuli (see Best et al., 1988). A significant increase in fixation during the postshift block relative to the preshift block is taken to indicate that infants detected the stimulus change. These data were entered into Language (3) × Trial Block (Preshift vs. Postshift) ANOVAs; test order was left out as a factor because preliminary analyses showed that it did not have any systematic effect on discrimination. Separate ANOVAs were conducted for each age to test the *a priori* predictions that 6-8 month olds would discriminate all three contrasts whereas 10-12 month olds would discriminate only the English and Zulu contrasts, and would fail on the Nthlakampx contrast. Given these predictions, simple effects tests of the Language × Trial Block interaction for each age were also carried out as planned comparisons.

6-8 month olds. This group's main effect for Trial Block, $F(1, 22) = 17.02$, $p < .002$, revealed significant recovery of fixation overall during the postshift trials ($M = 6.24$ s) relative to the preshift trials ($M = 2.56$ s), as expected. In addition, there was a significant Language effect, $F(2, 22) = 4.16$, $p < .03$. According to Tukey tests, this effect is attributable to significantly greater fixation for the English test ($M = 6.33$ s) than for the

Nthlakampx test (M = 3.27 s), $p < .05$, during the trials surrounding the shift. Although the Language × Trial Block interaction was not itself significant, a simple effects test on the interaction was conducted, as planned, to determine whether recovery of fixation during the initial test trial block was significant for each of the three contrasts. The results supported predictions. This age group showed significant recovery of fixation on the initial test trials for all three languages: English, $F_{(1,\ 11)} = 7.09, p < .025$; Zulu, $F_{(1,\ 11)} = 4.87, p < .05$; and Nthlakampx, $F_{(1,\ 11)} = 4.67, p = .05$. The simple effects tests also suggested that the main effect for Language could be traced primarily to fixation differences during the test trials, $F_{(2,\ 11)} = 2.97$, $p = .07$, rather than the preshift trials, *ns*. Postshift fixation was much higher for English (M = 9.19 s) than for Nthlakampx (M = 4.52 s), whereas preshift fixation was more nearly equivalent (M s = 3.5 vs. 2.03 s, respectively). Thus, while these infants showed significant discrimination on both tests, there is the suggestion of a mild language-specific effect in *degree* of discrimination for these two languages.

10-12 month olds. This age group also showed a significant Trial Block effect, $F_{(1,\ 11)} = 11.86, p < .004$, indicating overall discrimination (preshift M = 2.36 s; postshift M = 5.09 s). In contrast to the younger group, the Language effect was nonsignificant, while the Language × Trial Block interaction was marginally significant, $F_{(2,\ 22)} = 2.67$, $p = .09$. The planned simple effects test on this interaction revealed, as expected, that the older infants showed significant recovery of fixation only for English, $F_{(1,\ 11)} = 10.53$, $p < .008$, and for Zulu, $F_{(1,\ 11)} = 5.69, p < .04$, but not for Nthlakampx, *ns*. A comparison of the discrimination results for the two ages is shown in Figure 2.

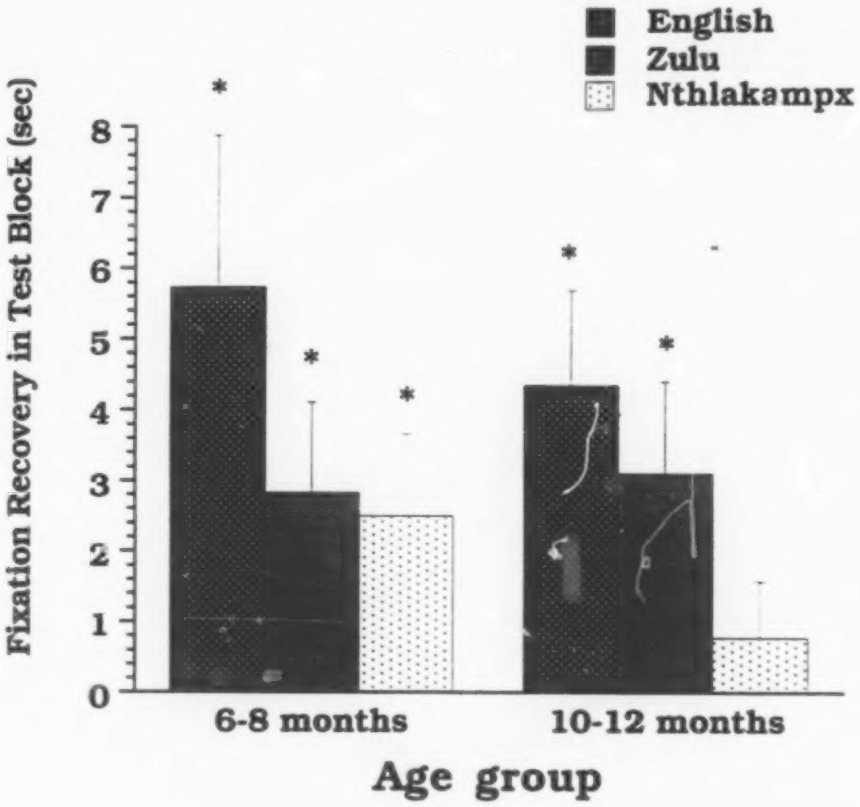


Figure 2. Recovery of fixation responses displayed as difference scores (preshift trial block - postshift trial block) with standard error bars, for each contrast at each age. Asterisks above the bars indicate those cases in which fixation during the first test block was significantly greater than fixation during the preshift trial block.

DISCUSSION

The findings strongly support the prediction of diverging developmental patterns for infants' perception of the two non-native consonant contrasts tested. The younger infants discriminated both non-native contrasts and, of course, the native English contrast. The older infants discriminated not only the English contrast but also the Zulu clicks, consistent with predictions for a NA contrast (Non-Assimilable) according to the Perceptual Assimilation Model (PAM), yet they failed to discriminate the Nthlakampx ejectives, consistent with predictions for SC assimilation (Single Category). The present study directly compared, in a single within-subjects investigation, two important but disparate cross-language speech perception findings with infants. Werker and Tees' (1984b) finding of developmental decline between 6-8 months and 10-12 months for English-learning infants' discrimination of the /k'-q'/ ejective distinction has now been replicated in an independent laboratory, and has been extended to a different methodological technique. We also replicated, in the same sample of infants, the earlier report that English-learning infants nonetheless continue to discriminate the /ǀa/-/ǁa/ click contrast even at 10-12 months, a finding which stands at odds with reports on perception of other non-native consonant contrasts by that age group.

The divergent developmental pattern for these two contrasts cannot be explained by any difference in phonetic exposure, since neither ejectives nor clicks occur as allophones of any English consonants. For this reason (as well as those provided in the Introduction), neither can the discrepant developmental trends for these two non-native contrasts be explained by differences in neural attrition due to differential phonetic exposure. Nor could the click vs. ejective discrimination difference at 10-12 months be caused by different degrees of auditory masking. The bursts of both the clicks and the ejectives, which appear to carry the primary information about both of these place of articulation contrasts, are separated from the following vowel by a silent gap that should sufficiently attenuate potential masking of the burst by the vowel.

Some might, alternatively, posit a difference in acoustic salience as an explanation of the difference in older infants' discrimination of the clicks versus the ejectives (see Burnham, 1986). However, as noted in Best et al. (1988), no objective criteria for salience have been proposed that are independent of the discrimination levels that salience is supposed to explain, thus the concept is tautological. In the present study, differential acoustic salience would be difficult to argue in any event, because these two particular non-native contrasts are quite similar in their acoustic properties, as described in the Introduction.

We suggest, instead, that the answer can be found in the development of perceptual ability to relate the physical, phonetic properties of speech to the more abstract phonological categories of the native language, as posited by PAM (Best, 1993, 1994, in press a, b; Best et al., 1988). Very young infants would not yet be expected to have determined the patterning of the native phonological system, and so it should not be surprising that they perceive the phonetic properties of both native and non-native segmental contrasts. However, at least by sometime in the second half of their first year, infants begin to recognize certain basic characteristics of the native phonological system, which in turn begins to influence their perception of non-native segmental contrasts. The present findings are consistent with the notion that discrimination of a non-native consonant contrast will be retained even at 10-12 months if the phonetic properties of that contrast place it outside the general patterning of phonetic properties within the native phonological system, that is, if the non-native consonants fit the definition of a NA contrast. Such was the case here for the Zulu click consonants. If on the other hand the non-native consonants fit the definition of a Single Category assimilation type, older infants would be expected to perceive them as phonetically equivalent and essentially indistinguishable, once they have begun to recognize certain basic properties of the native phonological system. The discrimination of the Nthlakampx ejectives by 6-8 month olds, but not by 10-12 month olds, fits this prediction as well. At the present time it is unclear, however, whether infants actually assimilate non-native consonants into native phonological categories in the same way that adults do, or even whether they have yet fully-specified phonological categories (see also Werker & Pegg, 1992). It would be reasonable to suppose that the answer to those two questions is "no," given that several years of further phonological and linguistic development must still take place after the first birthday before children have achieved adult-like levels of language competence. Additional research will be needed to address the issues of non-native phonetic assimilation and development of phonological categories in infants.

Interestingly, the data gathered in the present study with the conditioned fixation procedure revealed a hint of a language-specific effect in perception of a non-native contrast even at 6-8 months. That age showed more attention to English around the shift (primarily a difference in recovery of fixation at the stimulus shift) than to Nthlakampx velar-uvular ejectives, the latter being the same contrast that shows a significant decline in discrimination just a few months later. This language-specific bias was evident in the 6-8 month olds even though they nonetheless showed significant discrimination of the Nthlakampx ejectives. The pattern of language-specific change in perception of these ejectives differs strongly from the lack of developmental change in discrimination of the Zulu clicks. We suggest that this difference is due to perceived similarities between the ejectives and native voiceless stop consonants, but a lack of perceived similarities between the clicks and any English consonants. Again, however, further work will be needed to verify whether infants do indeed perceive phonetic similarity between ejectives and voiceless stops at the same place of articulation.

Given the suggestion by some (e.g., Burnham, 1968) that contrasts which remain easily discriminated, even without phonetic exposure may be acoustically salient, several characteristics of click consonants are of special note. If clicks are acoustically salient, they should presumably be easy to perceive and/or produce as *phonological* elements in the languages that employ them. In addition, they should be widespread across languages. But in fact, clicks are quite rare, being found only in the Khoisan languages of Africa, the language family that is the origin of the click consonants, and in those Bantu languages which borrowed the clicks from Khoisan-speaking groups over centuries of frequent interaction (Herbert, 1990). Linguists have posited a correlation between the ease of perceiving and/or producing a phonetic contrast and the commonness of that contrast across languages (e.g., see Lindblom, Krull, & Stark, 1993; Lindblom & Maddieson, 1988). Given the rareness of clicks across languages, they shouild be relatively difficult to discriminate *when perceived as phonological elements*. In keeping with this prediction, it is claimed that "to the untrained ear there is much confusion within the class" of click consonants (Herbert, 1990, p. 123) when non-click languages borrow words from a click language for example, the "borrowing" Bantu languages typically conflated a number of the original click distinctions found in the originating Khoisan languages, thus ending up with many fewer click distinctions than existed in the target language (Herbert, 1990). Additional evidence suggests articulatory difficulties with the production of clicks as phonological elements in languages. After the particular apical versus lateral click contrast we examined here, historical evidence indicates a strong tendency for those clicks to be conflated with others from the Khoisan languages. Specifically, the lateral click is currently disappearing in Zulu. The apical is next most likely to disappear in the adopting Bantu languages, such that the palatal has become the sole click in languages such as Sesotho (Herbert, 1990). In addition, anecdotal evidence (and a small amount of systematic observational evidence) on development indicates that the clicks develop relatively late in native-speaking children's productions (Jakobson, 1958; Louw, 1964). As was the case with click-borrowing languages, young children learning click languages show a strong preference to substitute palatal clicks in place of the lateral and apical clicks (Connelly, 1984; Herbert, 1983).

The perceptual findings from the present study are consistent with predictions based on PAM that there should be divergent developmental paths for perception of different types of non-native consonant contrasts. The present study supported the hypothesis that discrimination would show a developmental decline for a non-native contrast that adults are likely to assimilate into a single category in their native phonology. Complementary to that developmental pattern, support was also found for the prediction that discrimination would remain high across development for a contrast that adults fail to assimilate to any native categories, and therefore hear as nonspeech sounds. Further research will be needed, however, to corroborate PAM's predictions about other types of assimilation patterns, including TC (Two Category) and CG (Category Goodness difference) assimilation types.

REFERENCES

Aslin, R. N., & Pisoni, D. B. (1980). Some developmental processes in speech perception. In G. Yeni-Komshian, J. F. Kavanaugh, & C. A. Ferguson (Eds.), *Child phonology, Volume 1: Production*. New York: Academic Press.

Best, C. T. (1984). Discovering messages in the medium: Speech and the prelinguistic infant. In H. E. Fitzgerald, B. Lester, & M. Yogman (Eds.), *Advances in pediatric psychology* (Vol. 2, pp. 97-145). New York: Plenum.

Best, C. T. (1993). Emergence of language-specific constraints in perception of non-native speech: A window on early phonological development. In B. de Boysson-Bardies, S. de Schonen, P. Jusczyk, P. MacNeilage, & J. Morton (Eds.) *Developmental neurocognition: Speech and face processing in the first* *year of life* (pp. 289-304). Dordrecht, the Netherlands: Kluwer Academic Publishers.

Best, C. T. (1994). The emergence of native-language phonological influences in infants: A perceptual assimilation model. In J. Goodman & H. C. Nusbaum (Eds.), *The development of speech perception: The transition from speech sounds to spoken words.* (pp. 167-224). Cambridge MA: MIT Press.

Best, C. T. (in press a). Learning to perceive the Sound Pattern of English. To appear in C. Rovee-Collier & L. Lipsitt (Eds.), *Advances in infancy research* (Vol. 8). Norwood, NJ: Ablex Publishing Corporation.

Best, C. T. (in press b). A direct realist view of cross-language speech perception. To appear in W. Strange (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues in cross-language speech research.* Timonium, MD: York Press.

Best, C. T. McRoberts, G. W., & Sithole, N. N. (1988). The phonological basis of perceptual loss for non-native contrasts: Maintenance of discrimination among Zulu clicks by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance, 14*, 345-360.

Best, C. T., & Strange, W. (1992). Effects of phonological and phonetic factors on cross-language perception of approximants. *Journal of Phonetics, 20*, 305-330.

Burnham, D. K. (1986). Developmental loss of speech perception: Exposure to and experience with a first language. *Applied Psycholinguistics, 7*, 207-240.

Carney, A. E., Widin, G. P., & Viemeister, N. F. (1977). Noncategorical perception of stop consonants differing in VOT. *Journal of the Acoustical Society of America, 62*, 961-970.

Connelly, M. J. (1984). Basotho children's acquisition of morphology. Unpublished doctoral dissertation, University of Essex.

Diehl, R., & Kluender, K. (1989). On the objects of speech perception. *Ecological Psychology, 1*, 1-45.

Eilers, R. E., & Minifie, F. D. (1975). Fricative discrimination in early infancy. *Journal of Speech and Hearing Research, 18*, 158-167.

Eilers, R. E., Wilson, W. R., & Moore, J. M. (1977). Developmental changes in speech discrimination in infants. *Journal of Speech and Hearing Research, 20*, 766-780.

Eimas, P. D. (1975). Speech perception in early infancy. In L. B. Cohen & P. Salapatek (Eds.), *Infant perception: From sensation to cognition.* New York: Academic Press.

Eimas, P. D., & Miller, J. L. (1980). Discrimination of the information for manner of articulation by young infants. *Infant Behavior and Development, 3*, 367-375.

Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science, 171*, 303-306.

Flege, J. E. (1989). Chinese subjects' perception of the word-final English /t/-/d/ contrast: Before and after training. *Journal of the Acoustical Society of America, 86*, 1684-1697.

Flege, J. E. (in press). Second-language speech learning: Theory, findings, and problems. To appear in W. Strange (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues. Timonium*, MD: York Press.

Flege, J. E., & Eefting, W. (1987). The production and perception of English stops by Spanish speakers of English. *Journal of Phonetics, 15*, 67-83.

Herbert, R. K. (1983, July). Clicks in normal and delayed acquisition of Zulu. Paper presented at the *Tenth International Congress of Phonetic Sciences*, Utrecht, The Netherlands.

Herbert, R. K. (1990). The relative markedness of click sounds: Evidence from language change, acquisition, and avoidance. *Anthropological Linguistics, 32*, 120-138.

Jakobson, R. (1958). What can typological studies contribute to historical comparative linguistics? *Proceedings of the Eighth International Congress of Linguists* (pp.17-35). Oslo: Oslo University Press.

Jusczyk, P. W. (1993). Sometimes it pays to look back before you leap ahead. In B. de Boysson-Bardies, S. de Schonen, P. Jusczyk, P. MacNeilage, & J. Morton (Eds.) *Developmental neurocognition: Speech and face processing in the first year of life* (pp. 289-304). Dordrecht, the Netherlands: Kluwer Academic Publishers.

Jusczyk, P. W. (in press). Language acquisition: Speech sounds and the beginnings of phonology. In J. L. Miller & P. D. Eimas (Eds.), *Handbook of perception and cognition, Vol. 11: Speech, language, and communication.* Orlando FL: Academic Press.

Jusczyk, P. W., Kemler Nelson, D. G., Hirsh-Pasek, K., Kennedy, L., Woodward, A., Piwoz, J., (1992). Perception of acoustic correlates of major phrasal units by young infants. *Cognitive Psychology, 24*, 252-293.

Jusczyk, P. W., & Thompson, E. (1978). Perception of a phonetic contrast in multisyllabic utterances by 2-month-old infants. *Perception & Psychophysics, 23*, 105-109.

Kuhl, P. K. (1980). Perceptual constancy for speech-sound categories in early infancy. In G. H. Yeni-Komshian, J. F. Kavanaugh, & C. A. Ferguson (Eds.), *Child phonology: Vol. 2: Perception* (pp. 41-66). New York: Academic Press.

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science, 255*, 606-608.

Lasky, R. E., Syrdal-Lasky, A. & Klein, R. E. (1975). VOT discrimination by four to six and a half month old infants from Spanish environments. *Journal of Experimental Child Psychology, 20*, 215-225.

Lindblom, B., Krull, D., & Stark, J. (1993). Phonetic systems and phonological development. In B. de Boysson-Bardies, S. de Schonen, P. Jusczyk, P. MacNeilage, & J. Morton (Eds.), *Developmental neurocognition: Speech and face processing in the first year of life* (pp. 399-409). Dordrecht, the Netherlands: Kluwer Academic Publishers.

Lindblom, B., & Maddieson, I. (1988). Phonetic universals in consonant systems. In L. M. Hyman & C. N. Li (Eds.), *Language, speech, and mind.* London and New York: Routledge.

Lisker, L., & Abramson, A. S. (1970). The voicing dimension: Some experiments on comparative phonetics. *Proceedings of the 6th International Congress of Phonetic Sciences.* Prague: Academia.

Lively, S. E., Logan, J. S., & Pisoni, D. B. (1993). Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. *Journal of the Acoustical Society of America, 94*, 1242-1255.

Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. *Journal of the Acoustical Society of America, 89*, 874-886.

Louw, J. A. (1964). The consonantal phonemes of the lexical root in Zulu. *Afrika und Ubersee, 48*, 127-152.

MacKain, K.S. (1982). Assessing the role of experience on infants' speech discrimination. *Journal of Child language, 9*, 527-542.

MacKain, K. S., Best, C. T., & Strange, W. (1981). Categorical perception of English /r/ and /l/ by Japanese bilinguals. *Applied Psycholinguistics, 2*, 369-390.

Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N. Bertoncini, J., & Amiel-Tison, C. A. (1988). A precursor of language acquisition in young infants. *Cognition, 29*, 143-178.

Miller, C. L. (1983). Developmental changes in male-female voice classification by infants. *Infant Behavior and Development, 6*, 313-330.

Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A. M., Jenkins, J. J., & Fujimura, O. (1975). An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. *Perception & Psychophysics, 18*, 331-340.

Pisoni, D. B., Aslin, R. N., Perey, A. J., & Hennessey, B. L. (1982). Some effects of laboratory training on identification and discrimination of voicing contrasts in stop consonants. *Journal* *of Experimental Psychology: Human Perception and Performance, 8,* 297-314.

Polka, L. (1991). Cross-language speech perception in adults: Phonemic, phonetic, and acoustic contributions. *Journal of the Acoustical Society of America, 89,* 2961-2977.

Polka, L. (1992). Characterizing the influence of native experience on adult speech perception. *Perception & Psychophysics, 52,* 37-52.

Polka, L. (submitted). Linguistic influences in adult perception of non-native vowel contrasts.

Polka, L., and Werker, J. F. (1994). Developmental changes in perception of non-native vowel contrasts. *Journal of Experimental Psychology: Human Perception & Performance, 30,* 421-436

Streeter, L. A. (1976). Language perception of 2-month-old infants shows effects of both innate mechanisms and experience. *Nature, 259,* 39-41.

Swoboda, P. J., Kaas, J., Morse, P. A., & Leavitt, L. A. (1978). Memory factors in infant vowel discrimination in normal and at-risk infants. *Child Development, 49,* 332-339.

Swoboda, P. J., Morse, P. A., & Leavitt, L. A. (1976). Continuous vowel discrimination in normal and at-risk infants. *Child Development, 47,* 459-465.

Tees, R. C., & Werker, J. F. (1984). Perceptual flexibility: Maintenance or recovery of the ability to discriminate non-native speech sounds. *Canadian Journal of Psychology, 38,* 579-590.

Trehub, S. E. (1973). Infants' sensitivity to vowel and tonal contrasts. *Developmental Psychology, 9,* 91-96.

Trehub, S. E. (1976). The discrimination of foreign speech contrasts by adults and infants. *Child Development, 47,* 466-472.

Walley, A. C., Pisoni, D. B., & Aslin, R. N. (1981). The role of early experience in the development of speech perception. In R. N. Aslin, J. R. Alberts, & M. R. Petersen (Eds.), *Development of perception* (Vol. 1). New York NY: Academic Press.

Werker, J. F., Gilbert, J. H. V., Humphrey, K., & Tees, R. C. (1981). Developmental aspects of cross-language speech perception. *Child Development, 52,* 349-355.

Werker, J. F., & Lalonde, C. E. (1988). Cross-language speech perception: Initial capabilities and developmental change. *Developmental Psychology, 24,* 672-683.

Werker, J., & Logan, J. (1985). Cross-language evidence for three factors in speech perception. *Perception and Psychophysics, 37,* 35-44.

Werker, J. F., & Pegg, J. (1992) C. A. Ferguson, L. Menn, & C. Stoel-Gammon (Eds.) *Phonological development: Models, research, implications* (pp. 131-164). Timonium, MD: York Press.

Werker, J. F., & Tees, R. C. (1984a). Phonemic and phonetic factors in adult cross-language speech perception. *Journal of the Acoustical Society of America, 75,* 1866-1878.

Werker, J. F., & Tees, R. C. (1984b). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development, 7,* 49-63.

Williams, L. (1979). The modification of speech perception and production in second language learning. *Perception & Psychophysics, 26,* 95-104.

FOOTNOTES

[*] *Infant Behavior & Development,* in press.

[†] Also Wesleyan University, Department of Psychology, Middletown.

[‡] Stanford University, Department of Psychology, Palo Alto.

[†††] Wesleyan University, Department of Psychology, Middletown.

Haskins Laboratories Status Report on Speech Research
1993, SR-115/116, 69-97

Beyond Orthography and Phonology: Differences between Inflections and Derivations*

Laurie Beth Feldman†

The influence of morphological structure was investigated in two types of word recognition tasks with Serbian materials. Morphological structure included both inflectional and derivational formations and comparisons were controlled for word class and the orthographic and phonological similarity of forms. In Experiments 1, 2, and 3, the pattern of facilitation to target decision latencies was examined following morphologically-related primes in a repetition priming task. Although all morphologically-related primes facilitated targets relative to an unprimed condition, inflectionally-related primes produced significantly greater effects than did derivationally-related primes. In Experiments 4, 5, and 6 subjects were required to segment and shift an underlined portion from one word onto a second word and to name the result aloud. The shifted letter sequence was sometimes morphemic (e.g., the equivalent of ER in DRUMMER) and sometimes not (e.g., the equivalent of ER in SUMMER). Morphemic letter sequences were segmented and shifted more rapidly than their nonmorphemic controls when they were inflectional affixes but not when they were derivational affixes. These results indicate that (a) morphological effects cannot be ascribed to orthographic and phonological structure, (b) the constituent morphemic structure of a word contributes to word recognition and (c) morphemic structure is more transparent for inflectional than for derivational formations.

Morphology underlies the productivity of the word-formation process and a word's fit into the syntactic frame of a sentence. Linguists distinguish between two classes of morphological formations. Words that differ in their derivational affixes but share a base morpheme (e.g., CALCULATION, CALCULATOR) are generally considered to be different lexical items and to have different meanings. Words that differ in their inflectional affixes (e.g., CALCULATING, CALCULATED) but share a base morpheme are generally considered to be versions of the same word, with the particular version that appears in a sentence being determined by the syntax of the sentence. In general, inflectional formations are more productive, do not change word class membership relative to the base morpheme and are more constrained by syntax (Anderson, 1982) than are derivational formations. In addition, meanings of inflected forms tend to be compositional of the meaning of the base and affix morphemes, whereas meanings of derived forms are less often compositional. The present study examines how inflectional and derivational formations are processed.

Four principles of lexical storage have been proposed for words composed of several morphemes, that is morphologically-complex words. First, a principle of economic storage makes it appealing to represent complex forms in terms of a *base morpheme*. Accounts based on base morphemes are adequate for inflectional forms (e.g., Caramazza, Laudanna, & Romani, 1988) but are less plausible for derivations, in part, because (a) the formation rules for derivations are complex

---

The research reported here was conducted at the Laboratory for Experimental Psychology at the University of Belgrade and was supported by funds from National Institute of Child Health and Development Grant HD-01994 to Haskins Laboratories. Portions were presented to the November 1987 meeting of the Psychonomic Society. I wish to thank Darinka Andjelković who conducted the experiments and Dragana Barac-Cikoja and Petar Makara who helped prepare the materials. Valuable comments on earlier versions of this manuscript were provided by David Balota, Gary Dell, Carol Fowler, James Neely and several anonymous reviewers and I thank them all.

and there is no way to ascertain whether a particular form has been created and (b) the semantic contribution of the base morpheme to the meaning of the morphologically complex derivational form is unpredictable. Second, accounts based on the *stem* (base morpheme plus derivational affix, if any e.g., Burani & Laudanna, 1992) posit different representations for inflections and derivations. For example, in a lexical decision task where both items are formed around the same base morpheme, words with a derivational affix produce different patterns of facilitation between items relative to words with only an inflectional affix (Laudanna, Badecker, & Caramazza, 1992). Although it is likely that the lexical representation of inflected and derived forms differs, the relation between the two types of formations is underspecified. Third, morphologically complex words may be represented mentally as *whole forms*, without reference to their constituents (Butterworth, 1983). Fourth, Caramazza and his colleagues have proposed that *both base morpheme and whole word* are units for lexical access, that these alternatives are not mutually exclusive (e.g., Caramazza, Miceli, Silveri, & Laudanna, 1985), and that word frequency may play a key role (Caramazza et al., 1988).

The repetition priming paradigm (Stanners, Neiser, Hernon, & Hall, 1979), yielded evidence that morphological relationships constitute a principle of organization within the internal lexicon. The influence of morphological relatedness is assessed by comparing lexical decision latency or accuracy to the target preceded by a morphological relative to (a) a first presentation of the target word (i.e., no prime) and (b) an identical repetition of the target word. Sometimes the reduction in reaction times and errors that occurs with morphological relatives as primes is equivalent to the effect of an identical repetition (e.g., Fowler, Napps, & Feldman, 1985). Other times, decision latencies to targets following morphological relatives are reduced relative to first presentations but are slower than identical repetitions. The latter pattern is ambiguous. It has been interpreted as evidence of separate lexical entries (e.g., Stanners et al., 1979) and as evidence of interrelated entries (e.g., Fowler et al., 1985).

Facilitation due to morphological relatedness occurs in the lexical decision task across a variety of languages including Serbian (Feldman & Fowler, 1987), Hebrew (Bentin & Feldman, 1990), as well as English (Fowler et al., 1985; Feldman, 1992) and American Sign Language (Hanson & Feldman, 1989) and across a variety of conditions. Facilitation in repetition priming has been observed when prime and target are in either the same or different modalities (Fowler, et al., 1985; see also Kirsner, Milech, & Standen, 1983). Primes can be morphologically complex and targets can be morphologically simple or primes can be simple and targets complex (Feldman & Fowler, 1987; Schriefers, Friederici, & Graetz, 1992). The latter observation is significant because morphologically simple forms tend to be higher in frequency than morphologically complex forms. For complex targets, both derived and inflected formations show effects based on morphological relatedness (Fowler et al., 1985; Schriefers et al., 1992).

Several studies have tried to compare patterns of facilitation at long or at short lags for prime-target pairs related by inflection and by derivation. Differences in facilitation (ms) for targets following morphologically-related and unrelated primes are summarized in Table 1. As shown there, inflectional primes typically produce greater facilitation than derivational primes but the difference is often not statistically significant. For example, the words POTARONO and POTETE are related by inflection and the words POTATORE and POTETE are related by derivation. In a lexical decision task, both pairs produced faster latencies than unrelated pairs (Laudanna et al., 1992; Exp. 1). In these experiments, morphological relationship was defined on a single lexical item in Italian. Similarly, SPARIZIONE meaning *disappearance* is defined as a derivation whereas SPARIVANO meaning *they disappeared* is defined as an inflection. Only the latter slowed recognition of (morphologically-unrelated) SPARATI which is the past participle of *shot* and is formed from a different but homographic base morpheme (Laudanna et al., 1992; Exp. 3). In a repetition priming task with German materials (Schriefers et al., 1992), inflectional primes consisting of different inflected adjective forms produced greater facilitation than derivational primes consisting of abstract nouns formed from adjectives. Finally, in Hebrew (Feldman & Bentin, in press), morphological relationship was defined over the word pair because it is not always obvious which item is derived from which, but no differences between inflections and derivations were observed in the repetition priming task. In short, the pattern of results observed with various priming procedures indicates that differences between inflectional and derivational facilitation have appeared, but that they are often not reliably significant in separate comparisons.

Table 1. *Summary of facilitation for inflectional and derivational targets following identity, inflectional and derivational primes in immediate and long term priming tasks.*

	TYPE OF PRIME		INTERVAL	
IDENTITY	INFLECTIONAL	DERIVATIONAL[1]	DERIVATIONAL[2]	
STUDY:				
Stanners et al., 1979. Exp. 1[a]				long
166	181			
160	150			
140	131			
Stanners et al., 1979. Exp. 2[b]				long
84	49			
99	39			
Stanners et al., 1979. Exp. 3[c]				long
120		72		
118		32		
Fowler et al., 1985[d]				long
101	78			
42		47		
Feldman & Fowler, 1987. Exp. 1[e]				long
54	45			
Feldman & Fowler, 1987. Exp. 2[f]				long
90	74			
Feldman & Fowler, 1987. Exp. 3[g]				long
58	50			
Feldman & Bentin, 1992. Exp. 1[h]				long
68	60	59		
Schriefers, Friederici, & Graetz, 1992, Exp. 2[i]				long
108	99	50		
90		26	44	
Laudanna, Badecker, & Caramazza, 1992, Exp. 1[j]				short
	26	35		

[a] simple regular targets and inflected primes e .g., BURNS-BURN

[b] simple regular targets and irregular inflected primes e.g., HUNG-HANG

[c] simple regular targets and regular derived primes e.g., SELECTIVE-SELECT

[d] simple targets with sound change primes e.g., HEALTH-HEAL

[e] simple targets with regular inflected primes e.g., DINARA-DINAR

[f] inflected targets with simple and inflected primes e.g., DINAR(OM)-DINARA

[g] simple targets with inflected sound change targets PETKU/PETKOM-PETAK

[h] complex targets with complex primes e.g. NAFAL/NEFEL-NOFEL

[i] simple and complex targets with simple and complex primes e.g., ROTE/ROTLICH-ROT

[j] inflected targets e.g., RAPIVANO/RAPITORE-RAPIRE

The repetition priming results summarized above clearly demonstrate that under some conditions, morphological effects do arise in word recognition tasks. However, contrasts between the effects of inflections and derivations have not been compelling Some of the experiments included both inflectional and derivational forms but they did not explicitly compare these types of morphological formations. When planned comparisons between inflectionally- and derivationally-related prime-target pairs have been included, results have been equivocal. For example, whereas Stanners et al. (1979) reported significant differences in magnitude of facilitation for these two types of formations when they were regular only, Fowler et al. (1985) found no significant difference although small numerical differences typically were evident. Although these experiments with English materials included a comparison of facilitation with inflectional and derivational primes, this comparison is not without its problems. In English, inflectional formations tend to be more similar in form and meaning than are derivational formations (or alternatively, forms related by inflection share a stem as well as a base morpheme whereas forms related by derivation typically share only a stem). This observation is relevant because at short lags, orthographic overlap is sometimes reported to influence the pattern of facilitation in this and similar tasks (Emmorey, 1989; Napps & Fowler, 1987; Stolz & Feldman, in press). Moreover, the number of inflectional affixes for English is severely limited relative to the number of derivational affixes. These limitations impede a rigorous experimental comparison between inflectional and derivational formations with English materials.

By contrast, in Serbian it is possible to identify inflection-derivation pairs with only minimal differences in form and meaning. One such contrast entails agents and other nouns formed from verbs. For example, PEVAC, meaning *singer*, is formed from the verbal base morpheme PEV and the derivational affix AC. The same base morpheme appears in all present tense forms of the verb *to sing* including PEVA and PEVAM. Other sets of inflection-derivation pairs entail verb forms that share a base morpheme but differ in aspect (which reflects temporal properties of the verb). Perfective and imperfective aspect can be marked by the vowel of the suffix, by a prefix or by an infix. Although it is sometimes difficult to ascertain which is the derived form, it is well established that perfective and imperfective verbs in Serbian are derivationally related to each other. (Therefore, in the present study, derivation will be defined relative to a target.) Of course, each can be inflected to produce different verb forms. For both agent and aspect type of derivations, it is possible to identify inflectional forms with the same base morpheme so that the orthographic and phonemic overlap of primes with targets is matched across derivational and inflectional comparisons.

The first three of the present experiments were repetition priming experiments in which native speakers of Serbian performed a lexical decision task with Serbian materials. Targets were preceded by other forms that were either inflectionally or derivationally related to the target. Inflectional and derivational formations were matched for phonological and orthographic overlap with the target. In Experiment 1, targets such as PEVA (third person singular verb) were preceded an average of ten items earlier in the list by (a) an identical repetition, PEVA (b) the inflectionally-related prime, PEVAM (first person singular verb) or (c) the derivationally-related prime, PEVAC (nominative singular of agentive). Inflectional and derivational primes were matched with respect to orthographic and phonological similarity to the target but derivational forms did not preserve the word class of the target. In Experiment 2, prefixed or infixed imperfective verb targets in third person plural such as OBARE and GURNU were preceded by (a) an identical repetition, (b) an inflectionally-related prime, OBARIM or GURNEM (first person singular verbs), or (c) a derivationally-related prime, BARIM or GURAM (first person singular verbs) that differed in aspect. In Experiment 3, perfective targets such as NOSE (third person plural verbs) were preceded by (a) an identical repetition, (b) an inflectionally-related prime, NOSIM (first person singular verbs) or (c) a derivationally-related prime, NOSAM (imperfective first person singular verbs) where the last differed in aspect. Here, all primes and targets were verb forms and prime-target similarity was matched across one half of the inflectional and derivational primes. Using planned comparisons, target facilitation in lexical decision following inflectionally-related primes and derivationally-related primes was compared and facilitation following derivationally-related primes relative to first presentations was assessed.

In order to ascertain that the morphological effects observed in repetition priming were not specific to the lexical decision task, the effect of morphology was also investigated in a second experimental task. In Experiments 4, 5, and 6, subjects were required to segment and shift the final sequence of letters from a visually-presented source word to a target word and to name the new form aloud. Morphemic segments were compared with their phonemically- and orthographically-matched but nonmorphemic controls and both inflectional and derivational segments were examined. The structure of experimental materials for the present study is described in Table 2.

By linguistic accounts, the component structure of inflections is more transparent than that of derivations. The repetition priming task has proven itself to be sensitive to morphological relations between prime and target, but attempts to compare the patterns of facilitation in repetition priming for inflectionally- and derivationally-related prime-targets pairs have not yielded unambiguous results. This outcome may reflect the fact that in English, derivational affixes tend to be composed of more letters and to be semantically less compositional than are inflectional affixes. Experiment 1 was designed to compare these two types of morphological formations when effects of affix length are matched. In Serbian, it is easier to meet these constraints than in English because extensive families of words are formed from the same base morpheme.

Table 2. *The morphological constituents of morphologically simple and complex words in Serbian.*

	STEM			
WORD	BASE	SUFFIX: DERIVATION	SUFFIX: INFLECTION	MEANING
SERBIAN				
PEVAM	PEV		AM	I sing
PEVAČ	PEV	AČ		singer
NOSIM	NOS		IM	I carry (perfective)
NOSAM	NOS		AM	I carry (imperfective)
PRESOM	PRES		OM	press (instrumental)
PRELOM	PRELOM		fracture	
CEVI	CEV		I	pipes (nominative plural)
CEDI	CED		I	he wrings
JEDEM	JED		EM	I eat
BEDEM	BEDEM			embankment
ZIDAR	ZID	AR		brick layer
KADAR	KADAR			sequence
BAJAM	BAJ	AM		I do magic
SAJAM	SAJAM			fair
BAŠTICA	BAŠT	ICA		garden (diminutive)
KOŠTICA	KOŠTICA			pit

Methods

Subjects. Twenty-seven first year students from the Department of Psychology at the University of Belgrade participated in Experiment 1. All were native speakers of Serbian. All had vision that was normal or corrected to normal and had prior experience in reaction-time studies.

Stimulus materials. Twenty-seven Serbian word triples were selected. Fourteen consisted of a noun target in nominative case with an inflectionally-related form in instrumental case and a derivationally-related verb form. For example, the nominative target BROD, meaning *boat,* was paired with its instrumental BRODU and with BRODI, the third person singular of the verb meaning *to sail* which is derivationally-related to BROD, the target. The remaining thirteen triples consisted of verb targets in one of three singular person forms with another inflected form of that same verb and with the agentive derived from that verb. For example, the target PEVA, meaning *he sings,* was paired with PEVAM, meaning *I sing,* and PEVAC, meaning *singer*. All words were highly familiar, contained between three and seven letters, and were printed in Roman script. They are listed in Appendix 1. Twenty-seven orthographically and phonemically regular pseudowords were generated by changing one or two letters (vowel with vowel or consonant with consonant) in bases of other real words. Triplets were generated for these pseudowords in a fashion analogous to that for words (i.e., affixes were real).

A member of each morphologically-related word (and pseudoword) triple appeared once as a target and once as a prime. In the identity condition, the same form occurred twice. In the inflectionally-related condition, the prime was another inflected form of the target and it necessarily preserved word class. In the derivationally-related condition, the prime was a verbal form for noun targets and a noun form for verb targets. Inflectionally and derivationally related primes were each one or two letters longer than the target and within a pair, overlap was perfectly matched phonemically as well as orthographically. Finally, the full target was contained within the inflectionally and derivationally related primes. For example, both BRODI and BRODU are each one letter longer than and include the target BROD. Primes and targets were separated by an average of 10 intervening items with a range was of 8 to 12 items.

Procedure. Individually tested subjects performed a lexical decision task. On each trial, a visual fixation signal accompanied by an auditory signal appeared for 200 ms then a target letter string printed in upper case was presented for 750 ms. As each target letter string appeared on the CRT of an Apple II computer, the subject pressed a telegraph key with both hands to indicate whether or not it was a word. A press of the farther key signaled "yes" and the closer key, "no". Reaction time was measured from the onset of the letter string. The interval between subject's response and the onset of the next experimental trial was 2000 ms.

Design. For Experiment 1, three test orders each containing 114 items were created. Half of the items were words and half pseudowords. Fifty-four items were primes and fifty-four items were targets. Words and pseudowords were presented equally often as primes and as targets. In addition, there were six filler items introduced to maintain the requisite lags. Each test order included nine tokens of each of the three types of primes (viz., identity, inflectional, derivational) and across test orders, each target was preceded by all three types of prime. Subjects were randomly assigned to one of the three test orders and a practice list of ten items preceded each experimental list.

Results and Discussion

Errors and extreme response times (greater than 2 SD or less than -2 SD from each subject's mean) were eliminated from all reaction time analyses. Accordingly, about 4% of all responses were eliminated. Table 3 summarizes the mean recognition times over subjects for target words and pseudowords preceded by identity, inflectional and derivational primes and for the first presentation of those same words as a prime.

Analyses of variance were performed on target latencies for words and pseudowords using subjects *(F1)* and items *(F2)* as random variables. The analysis included the first presentation of the target as the no prime condition, targets preceded by themselves as the identity condition, targets preceded by an inflected form and targets preceded by a derived form. For words, there was a significant effect of type of prime on target latencies [$F1(3,78) = 15.66$ $MS_e = 641$, $p < .001$; $F2(3,78) = 8.13$, $MS_e = 1597$, $p < .001$] but the effect of prime with the error measure missed significance [$F1(3,78) = 2.66$, $MS_e = 46$, $p < .054$; $F2(3,78) = 0.88$]. The results of planned comparisons on decision latencies indicated that facilitation from derivationally-related primes was significantly weaker than facilitation from inflectionally-related primes [$F1(1,26) = 6.58$, $MS_e = 383$, $p < .016$; $F2(1,26) = 3.1$, $MS_e = 4950$, $p < .08$] and significantly different from the no prime condition [$F1(1,26) = 6.14$, $MS_e = 804$, $p < .02$; $F2(1,26) = 4.05$, $MS_e = 6468$, $p < .05$]. Target latencies following derivational primes tend to be slower than target latencies following inflectional primes. For pseudoword targets, the effect of type of prime was significant for neither reaction times nor errors.

Table 3. *Mean lexical decision latencies and errors for targets on their first presentation, or when preceded by identity, inflectionally- and derivationally-related primes in Experiment 1.*

	TYPE OF PRIME							
	NONE		IDENTITY		INFLECTIONAL		DERIVATIONAL	
	PEVA ’		PEVA PEVA		PEVAM PEVA		PEVAČ PEVA	
	RT	ERR	RT	ERR	RT	ERR	RT	ERR
WORDS								
	569	(8)	524	(8)	536	(9)	550	(12)
FACILITATION			45	(0)	33	(-1)	19	(-4)
PSEUDOWORDS								
	664	(8)	666	(8)	663	(9)	644	(7)
FACILITATION			-2	(0)	1	(-1)	20	(1)

Facilitation was assessed by examining differences in latencies (and errors) to targets preceded by a prime and to first presentations of targets. Consequently, prime presentations necessarily occurred earlier in the list than did targets. Because there is evidence that latencies get faster as subjects proceed through the list, and because facilitation following derivations tended to be weak relative to facilitation following inflections, it is important to determine whether or not facilitation from derivations was correlated with serial position of the prime. In Experiment 1, the correlation between serial position of the prime and the difference between latencies for first presentations and latencies following derivational primes was $r = -.048$. Therefore, the magnitude of facilitation was not distorted by the no prime baseline. Note, however, that any potential baseline problem is not relevant when comparing facilitation following inflectional and derivational primes because position of the target (and the target item) were identical.

The present experiment with Serbian materials replicates previous findings in the same language (Feldman & Moskovljević, 1987; Feldman & Fowler, 1987) as well as other languages (Fowler et al., 1985; Bentin & Feldman, 1990). Specifically, relative to a no prime condition, morphologically-related word forms facilitated each other at lags that average 10 intervening items but pseudoword analogs did not. In summary, facilitation was observed in the pattern of target latencies for all types of morphologically-related primes and the amount of facilitation varied by type of prime. It is interesting to note that although identity and inflectional primes tended to yield statistically equivalent facilitation in earlier studies (e.g., Feldman & Fowler, 1987), under some circumstances derivations have been observed to produce facilitation that was significantly reduced relative to the identity condition (Feldman & Moskovljević, 1987; Exp. 2; Schriefers et al., 1992). Nevertheless, no published experiment with Serbian materials included, or even permitted, a direct comparison between inflectional and derivational types of primes.

The present study extends previous repetition priming results in Serbian by contrasting two types of morphological formations while tightly controlling their similarity. With phonemic and orthographic overlap equated between inflectionally- and derivationally-related prime forms, there was evidence of enhanced facilitation for targets following inflectionally-related primes relative to derivationally-related primes. This distinction can be represented in the lexicon. Perhaps the linkage between whole word forms that share a base morpheme is stronger (or the internal coherence of their constituents is weaker) for inflectionally-related forms than for derivationally-related forms.

Unfortunately, the composition of experimental materials in the present experiment is consistent with another account. In Experiment 1, all derivational formations differed in word class from their morphologically-related target whereas all inflectional formations (necessarily) preserved word class. Specifically, translations of agentives such as *singer* primed verb targets such as *he sings* and verb forms such as *he sails* primed derived noun targets such as *boat*. While such changes are, in fact, characteristic of derivational processes in all languages, they make an unequivocal interpretation of the contrast between inflectional and derivational pairs more difficult. It is important to note that although, in the repetition priming task, no effects of semantic similarity have been reported with visually presented relatives and lags of 10 items (Bentin & Feldman, 1990; Napps, 1989) or with auditorily presented materials presented successively (Emmorey, 1989; but see Radeau, 1983; Slowiaczek, 1994), it is nevertheless possible that derivations are semantically more distinct from their targets than are inflections and that semantic similarity can, under some circumstances, contribute to the pattern of facilitation. Accordingly, Experiment 2 entailed a comparison of the pattern of facilitation with Serbian inflections and derivations that a) consistently preserved word class, b) were semantically quite close in meaning and c) were constructed with attention to their orthographic similarity to the target.

EXPERIMENT 2

Inflectional affixes tend to alter the meaning of the base morpheme in predictable ways (Aronoff, 1976) whereas the effect of derivational affixes is less consistent. Consequently, inflectional formations tend to be similar in form and meaning to other forms that share a base morpheme (and stem) and differ with respect to inflectional affix whereas derivational formations tend to differ in form and meaning from other forms that share a base morpheme and differ with respect to derivational affixes (and stem). In Serbian it is possible to identify inflection-derivation pairs with only minimal differences in meaning and form.

One such contrast entails verbs that differ with respect to aspect. Generally stated, aspect reflects the temporal properties of the verb. These include inceptive forms of stative verbs and iterative forms of verbs that describe discrete events.

All the experimental materials for Experiment 2 were verb forms. Targets were preceded by identity, inflectionally- or derivationally-related primes. Inflected primes were other forms of the same verbs that differed in person. Derived primes were forms of lexically-distinct verbs composed from the same base morpheme that differed in aspect and person from the target word. The manipulation on derivation alternated perfective and imperfective forms. Semantically, this distinction is relatively minor entailing contrasts between semantic notions such as completed and progressive actions in HE SAT DOWN and HE WAS SITTING or between events and states such as HE RECOGNIZES and HE KNOWS. (Note that progressivity is grammaticalized in English whereas stativity is lexicalized (Lyons, 1977)). It is important to underscore, however, that in Serbian, unlike English, the perfective and imperfective forms of the verbs included in the present study are considered distinct lexical entries.

It is relevant to note that there is no consensus about the morphological status of aspectual formations either across languages or across theorists (compare Anderson, 1982 with Bybee, 1985). In the present study, it is assumed that aspect is a derivational process. It is restricted by its meaning to a particular semantic class of Serbian verbs (Partridge, 1964 in Bybee 1985). Moreover, it was also always the case that two distinct verbal entries existed in the dictionary. Note however, that these formations do not change word class as is typical of derivational formations. In Experiment 2, aspect was marked by the addition of either a prefix or an infix to the base morpheme. Consequently, forms related by inflection shared both a base morpheme and a stem (base morpheme plus derivational affix) whereas forms related by derivation shared a base morpheme but differed with respect to stem. For example, the words OBARIM and BARIM are both formed from the base BAR and the inflectional affix IM. They differ with respect to the presence of a prefix which is part of the stem. Accordingly, the stems are OBAR and BAR, respectively.

The outcome of Experiment 1 indicated that with controls for orthographic overlap, the lexical representation of morphological relatedness by inflection and derivation differed. If this outcome reflects type of morphological relation as distinguished from effects of preserving or altering word class, then consistent with the results of Experiment 1, in Experiment 2 facilitation from primes that are inflectionally-related to their targets should be greater than from primes that are derivationally-related. Of course, the absence of a difference is ambiguous. It could indicate that the effect observed in Experiment 1 does reflect changes in word class between prime and target. Alternatively, it could indicate that aspect in Serbian is not a derivational relationship but rather, a less general inflectional relationship.

If, as sometimes claimed (e.g., Taft & Forster, 1975; Bergman, Hudson, & Eling, 1988), prefixes but not other affixes are stripped from the base before lexical access is attempted, then the pattern for prefixed primes should differ from that of infixed primes. Alternatively, if activation in repetition priming is based on the stem (base plus derivational affix) rather than the base alone as sometimes claimed (Burani & Laudanna, 1992) then infixed forms should show a pattern similar to that of prefixed forms. Because inflections shared both base and stem whereas derivations shared a base morpheme only, derivations should produce weaker facilitation than inflections whenever the stem and base morpheme differ.

In summary, as in Experiment 1, patterns of facilitation for primes related by inflection and by derivation are examined in Experiment 2. Both inflectional and derivational primes always included the full base morpheme and their inflectional affixes were matched for letter length. In contrast to Experiment 1, in which orthographic and phonological overlap was perfectly matched but word class differed between derivational but not inflectional primes, in Experiment 2, the presence of a prefix or an infix rendered inflectional primes more similar to their targets than derivational primes (that included no affix) but all were verb forms.

Methods

Subjects. Thirty-six first year students similar in characteristics to those of Experiment 1, participated in Experiment 2. None had participated in Experiment 1.

Stimulus materials. Forty-eight Serbian word triples were selected. Each included three verb forms: a target verb, a prime that was inflectionally related and a prime that was derivationally related to the target. Targets consisted of present-tense verb forms in the third person plural. Each was composed of a base morpheme and an aspectual affix. Inflected forms were first person singular of those same verbs. Derived forms were first person singular of different verbs formed from the same base morpheme without an aspectual affix. (These forms are designated as derived because they are related by derivation to the target.) Inflectional and derivational primes were always presented in the same person and number. Items are listed in Appendix 1.

Typically, the target and inflected prime were imperfective forms and the derived prime was perfective. They were all formed from the same base morpheme but, because of the addition of an affix, they differed with respect to their stems. Derivation was defined relative to the target rather than on an isolated word. Structurally, all members of a triple were composed of the same base morpheme but differed with respect to the presence of an affix, either prefix or infix. For example, perfective forms of the base morpheme BAR, meaning *cook*, included BARIM, BARIS, BARI...BARE whereas imperfective forms such as OBARIM, OBARIS, OBARI...BARE include the prefix O. Other than the prefix or infix, the orthographic and phonemic overlap of primes and their morphologically-related targets was perfectly controlled by selecting third person plural forms ending in E as targets and necessarily as identity primes (e.g., OBARE), forms ending in IM (e.g., OBARIM) as inflectionally-related primes and verbs differing in aspect (e.g., BARIM) as the derivationally-related primes. Perfective forms of the base morpheme GUR, meaning *push*, include GURAM, GURAS, GURA...GURAJU whereas imperfective forms such as GURNEM, GURNES, GURNE...GURNU include the infix N. For infixed relatives, targets and identity primes ending in U (e.g., GURNU), inflectional primes ending in EM (e.g., GURNEM) and derivational primes ending in AM (e.g., GURAM) were presented where inflectional and derivational primes were always in the same person and number. In summary, the orthographic and phonological similarity of both inflectional and derivational primes was matched to the target so that both included the full base morpheme although due to the prefix or infix, overall overlap for inflectional forms was slightly greater than that for derivational forms.

Pseudoword triples were created by substituting vowels or consonants within other base morphemes in order to create meaningless bases that were orthographically legal. To these, real inflected affixes were appended in order to create sets of pseudowords that differed only with respect to affix. The distribution of pseudoword affixes was matched to those for words. Pseudoword targets were preceded by identity, inflectionally- (or derivationally-) related pseudoword primes or by a real word prime. The value in including a word prime with a pseudoword target was to examine whether facilitation in repetition priming extends to strings without lexical status.

Three test orders were created. Each contained 200 items and included equal numbers of word and pseudoword targets preceded an average of ten items earlier in the list by a morphologically-related prime. In each test order, eight tokens for each of the three types of prime were presented. Across the three test orders, each word or pseudoword target was preceded by all three types of morphologically related primes. In contrast to previous repetition priming studies, here pseudoword targets were preceded 33% of the time by a word prime formed from the same base morpheme.

Procedure. The procedure was identical to that of Experiment 1.

Results and Discussion

Mean decision latencies (for responses less than 2 SD or greater than -2 SD from each subject's mean) and error rates in Experiment 2 are summarized in Table 3. Errors and outliers accounted for approximately 6% of all responses. An analysis of lexical decision latencies for words revealed a significant effect of type of prime [$F1(3,105) = 45.06$, $MS_e = 1726$, $p < .001$; $F2(3,138) = 27.73$, $MS_e = 2865$, $p < .0001$]. Effects of affix type were significant [$F1(1,35) = 46.51$, $MS_e = 1321$, $p < .001$; $F2(1,46) = 5.42$, $MS_e = 12712$, $p < .02$]. The interaction of type and overlap was significant in the subjects but not in the items analysis [$F1(3,105) = 4.59$, $MS_e = 2101$, $p < .005$]. Planned comparisons indicated that inflectional primes produced faster target latencies than did derivational primes both for prefixed targets [$F1(1,35) = 4.61$, $MS_e = 2596$, $p < .04$; $F2(1,23) = 3.88$, $MS_e = 12096$, $p < .053$] and for infixed targets, [$F1(1,35) = 8.79$, $MS_e = 1759$ $p < .005$; $F2(1,23) = 6.79$, $MS_e = 17749$, $p < .01$]. Target latencies following derivational primes were significantly faster than first presentations latencies in the prefixed [$F1(1,35) = 16.23$, $MS_e = 2660$, $p < .001$; $F2(1,23) = 11.26$, $MS_e = 35101$, $p < .001$] but not in the infixed condition [$F1(1,35) = 1.96$, $MS_e = 2883$, $p < .17$; $F2(1,23) = 4.74$, $MS_e = 12384$, $p < .03$]. Correlations between serial position of the prime and the magnitude of facilitation for targets (following derivational primes relative to the no prime condition) were not significant [$r = -.154$] therefore it is unlikely that the magnitude of facilitation was significantly overestimated by the no prime baseline.

The analysis of error scores revealed a significant effect of type of prime [$F1(3,105) = 17.01$, $MS_e = 85$, $p <.001$; $F2(3, 138) = 5.84$, $MS_e = 1.8$, $p < .001$]. Effect of affix type were significant in the subjects analysis only [$F1(1,35) = 32.89$, $MS_e = 79$, $p < .001$; $F2(1,46) = 2.99$, $MS_e = 8.2$, $p < .09$]. The interaction of type and overlap was significant in the analysis by subjects only [$F1(3,105) = 3.30$, $MS_e = 85$, $p < .023$; $F2(3,138) = 1.45$, $MS_e = 1.8$, $p < .23$]. Mean decision latencies and errors for pseudowords are included in Table 4. They indicate no effect of lexical status of the prime on pseudoword target latencies.

The magnitude of facilitation was significantly greater for prime-target pairs related by inflection than for pairs related by derivation. Thus, facilitation in repetition priming was once again sensitive to type of morphological relation when word class and formal properties of the affixes were controlled. Assuming an appropriate baseline, derivational primes produced significant facilitation relative to the no prime condition for prefixes although not (statistically) for infixes. More importantly, derivational primes produced significantly reduced facilitation relative to the inflectional primes. The present results replicate the general pattern of morphological relatedness in the repetition priming task including the different patterns of facilitation following inflectional and derivational primes that was observed in Experiment 1.

Table 4. *Mean lexical decision latencies (and errors) for targets on their first presentation, or when preceded by identity, inflectionally- and derivationally-related primes in Experiment 2.*

	TYPE OF PRIME							
	NONE		IDENTITY		INFLECTIONAL		DERIVATIONAL	
PREFIXED WORDS								
	OBARE		OBARE OBARE		OBARIM OBARE		BARIM OBARE	
	RT	ERR	RT	ERR	RT	ERR	RT	ERR
	675	(9)	57	(4)	600	(5)	626	(8)
FACILITATION			102	(5)	75	(4)	49	(1)
INFIXED WORDS								
	GURNU		GURNU GURNU		GURNEM GURNU		GURAM GURNU	
	675	(21)	629	(8)	628	(7)	658	(14)
FACILITATION			46	(13)	47	(14)	17	(7)
PSEUDOWORDS								
	678	(7)	672	(5)	668	(6)	665	(6)
FACILITATION			6	(2)	10	(1)	13	(1)

EXPERIMENT 3

The natural confound between inflections and derivations noted above was eliminated in the third experiment. Specifically, forms related by inflection tend to be more similar in terms of orthography and phonology than forms related by derivation. This is because derived forms share a base morpheme but differ with respect to derivational affix and therefore stem whereas inflected forms share both their base morpheme and their stem. The materials for Experiment 3 consisted of another set of verbs related by aspect. In each instance, two entries were formed around the same base morpheme; however, they differed with respect to the set of inflectional affixes each required. That is, many items shared both their base morpheme and their stem and they differed only with respect to their thematic vowel (Scalise, 1984). If differences between facilitation by inflection and derivation are observed with the materials of Experiment 3, they cannot be attributed to orthographic overlap or to repetition of the base morpheme but not the stem.

Methods

Subjects. Thirty-six first year students similar in characteristics to those of the first two experiments participated in Experiment 3. None had participated in Experiments 1 or 2.

Stimulus materials. Twenty-six word triples in Serbian were selected. Each included three verb forms: a target verb, a prime that was inflectionally-related, and a prime that was derivationally-related to the target. Targets consisted of present tense verb forms in the first or third person plural. Inflected forms were first person singular of those same verbs. Derived forms were first person singular of different verbs formed from the same base morpheme that differed in the temporal qualities of the action they conveyed. Inflectional and derivational primes were always presented in the same person and number. Items are listed in Appendix 1.

The orthographic and phonemic overlap of primes and their morphologically-related targets was carefully controlled and two patterns were included. Structurally, all members of a triple in the *matched* pattern were verbs constructed from the same base morpheme and stem but they differed with respect to the (thematic) vowel around which the inflectional affix was formed. For example, in one pattern, perfective forms of the base morpheme NOS meaning *carry* are generally formed around the vowel I such as NOSIM, NOSIS NOSI...(but) NOSE whereas imperfective forms are generally formed around A such as NOSAM, NOSAS NOSA...NOSAJU. Forms ending in E served as targets and necessarily as identity primes (e.g., NOSE), forms ending in IM or EM (e.g., NOSIM) served as inflectionally-related primes, and verbs differing in aspect (e.g., NOSAM) served as the derivationally-related primes. Thirteen such pairs were selected. Thirteen pairs followed a second *unmatched* pattern in which the inflectionally-related prime overlapped by one or two letters more than did the derivationally-related prime. For example, forms ending in AMO (e.g., NAZIVAMO) served as targets and as identity primes, forms ending in AM (e.g., NAZIVAM) served as inflectional primes and forms ending in EM (e.g., NAZOVEM) served as derivational primes. Note that for these triples, inflectional primes preserved both the I and A vowels of the target whereas the derivational primes did not.

Pseudoword triples were created by substituting vowels or consonants within other base morphemes in order to create meaningless bases that were orthographically legal. To these, real inflected affixes were appended in order to create sets of pseudowords that differed only with respect to affix. As with words, pseudoword targets were preceded by identity, inflected and derivationally-related primes. Inflected and derived forms consisted of a nonsense base morpheme with a legal affix. The distribution of pseudoword affixes was matched to those for words.

In summary, as in Experiments 1 and 2, both inflectional and derivational primes always included the full base form and their affixes were matched for letter length. In contrast to Experiment 1, in which orthographic and phonological overlap was perfectly matched but word class differed between prime and target, derivational as well as inflectional primes in Experiments 2 and 3 preserved word class of the target. In contrast to Experiment 2, in which orthographic and phonological overlap between inflectional and derivations primes was not perfectly matched, in Experiment 3 matched and mismatched overlap was systematically manipulated. In the condition in which orthographic and phonological overlap were mismatched, inflectional primes were more similar to their targets than were derivational primes because the inflectional primes (i.e., those ending in AM) preserved the vowels of the target form whereas none of the derivational primes did. In the condition in which orthographic and phonological overlap were matched, inflectional and derivational primes were equally similar to their targets.

Three test orders were created. Each contained 114 items and included equal numbers of word and pseudoword targets preceded an average of ten items earlier in the list by a morphologically-related prime. In each test order, four or five tokens of each of the three types of matched and unmatched primes were presented. Across the three test orders, each word or pseudoword was preceded by all three types of morphologically related primes.

Procedure. The procedures were identical to those of Experiments 1 and 2.

Results and Discussion

Mean decision latencies (for responses less than 2 SD or greater than -2 SD from each subject's mean) and error rates in Experiment 3 are summarized in Table 5. Accordingly, approximately 5% of responses were eliminated. An analysis of lexical decision latencies for words revealed a significant effect of type of prime [$F1(3,105) = 23.40$, $MS_e = 2598$, $p < .001$; $F2(3,72) = 11.17$, $MS_e = 2396$, $p < .0001$]. Effects of orthographic and phonological overlap (match) were significant in the analysis by subjects [$F1(1,35) = 21.14$, $MS_e = 3607$, $p < .001$] but not in the analysis by items [$F2(1,24) = 2.54$; $MS_e = 8618l$; $p < .12$]. Importantly, the interaction of type and match did not approach significance in either analysis. Nevertheless, comparisons between inflections and derivations were examined separately for matched and mismatched items. Target latencies following orthographically-matched inflectional primes were faster than target latencies following derivational primes [$F1(1,35) = 6.26$, $MS_e = 2197$, $p < .014$; $F2(1,12) = 4.73$, $MS_e = 1150$, $p < .05$]. However this pattern missed significance for mismatched primes [$F1(1,35) = 2.06$, $MS_e = 2197$, $p < .15$; $F2(1,12) = 1.23$]. Finally, latencies for targets following derivational primes were significantly faster than for first presentations when overlap was mismatched [$F1(1,105) = 28.58$ $MS_e = 2197$, $p < .001$; $F2(1,12) = 8.81$, $MS_e = 2814$, $p < .01$] and in the subjects analysis, when overlap was matched [$F1(1,105) = 6.97$, $MS_e = 2197$, $p < .01$; $F2(1,12) = 1.62$, $MS_e = 4240$, $p < .23$]. The correlation between serial position of the prime and magnitude of facilitation for targets following derivational primes relative to the no prime condition was positive and nonsignificant [$r = .11$]. In conjunction with previous experiments, this finding supports the appropriateness of the no prime baseline.

The analysis of error scores revealed a significant effect of type of prime [$F1(3,105) = 4.60$, $MS_e = 78$, $p < .005$; $F2(3,75) = 3.74$, $MS_e = .61$, $p < .02$]. For pseudowords, neither decision latencies nor accuracy revealed an effect of prime. Mean decision latencies and errors for pseudowords are included in Table 5.

Table 5. *Mean decision and naming latencies (and errors) for targets on their first presentation, preceded by identity, inflectionally- and derivationally-related primes in Experiment 3.*

	TYPE OF PRIME							
	NONE		IDENTITY		INFLECTIONAL		DERIVATIONAL	
	NOSE		NOSE NOSE		NOSIM NOSE		NOSAM NOSE	
	RT	ERR	RT	ERR	RT	ERR	RT	ERR
MATCHED WORDS								
	630	(10)	577	(2)	573	(3)	600	(4)
FACILITATION			53	(8)	57	(7)	30	(6)
MISMATCHED WORDS								
	676	(4)	617	2)	601	(3)	617	(2)
FACILITATION			59	(2)	75	(1)	59	(2)
	NONE		IDENTITY		PSEUDOWORD		WORD	
PSEUDOWORDS								
	674	(8)	671	(9)	665	(8)	663	(9)
FACILITATION			3	(-1)	9	(0)	11	(-1)

The pattern of target latencies indicated that identical repetition and inflectional primes both produced significant and equivalent facilitation. Matched derivational primes produced significantly reduced facilitation relative to the inflectional condition and significant facilitation relative to the no prime condition. The present results replicate previously observed effects of morphological relatedness in the repetition priming task and extend those results by revealing a significant distinction between the effect on targets of inflectional and derivational primes that share both their stem and their base morpheme.

Effects of orthographic and phonologic overlap between prime and target on the pattern of facilitation across prime types were systematically examined because inflectional relatives tend to be more similar than derivational relatives. Matched and mismatched overlap never interacted with type of prime although planned comparisons indicated that the difference between targets preceded by inflections and by derivations was statistically more reliable for matched pairs than for mismatched pairs. This pattern is not anticipated if differences in magnitude of facilitation between inflectionally- and derivationally-related primes reflects extent of orthographic overlap with the target. Moreover, because the semantic differences between inflectional and derivational relatives was small, it cannot readily be attributed to greater semantic overlap for inflections relative to derivations.

The materials selected for Experiment 3 are unique in that for many items inflectional and derivational relatives were both formed around the same base morpheme and differed only with respect to the vowel from which the inflectional affixes were formed. Because no derivational affix was introduced, relatives shared both their base morpheme and their stem. Thus, the results of Experiment 3 indicate that the difference between inflections and derivations in the repetition priming task cannot be attributed to greater facilitation for stems than for bases.

In the present study, morphological relatives produced facilitation to target decision latencies in the repetition priming task but the interpretation of these lexical decision results is not straightforward. It has been suggested that the results obtained with this task may reflect binary decision processes that are specific to this task (Balota & Chumbley, 1984) or alternatively that expectancy and post-lexical mechanisms are involved as well as lexical activation (Neely, 1991). Obviously, it is important to provide converging evidence from other experimental tasks for the contribution of morphology to word recognition. In the three remaining experiments, morphological effects are investigated in a new experimental task.

EXPERIMENT 4

The outcome of the first three experiments using the repetition priming paradigm suggested processing differences between inflectional and derivational formations. Another source of evidence for the role of morphology in lexical processing derives from the pattern of errors observed in the production of spontaneous speech (Cutler, 1980; Dell, 1986; Fromkin, 1973; Garrett, 1980, 1982; Stemberger, 1985). One prevalent type of error entails the reordering of morphemic elements so that the stem or affix of a word migrates from the intended word to another site. The pattern for stems and affixes tend to differ (Garrett, 1976). Although there are confounded prosodic differences, this observation has been interpreted as evidence that the base morpheme and inflectional components of a morphologically complex word are separable. Moreover, when word final elements are misordered, those that are morphemic are more likely to shift than are phonemically equivalent but nonmorphemic segments (Stemberger, 1984) and this difference cannot be attributed to frequency differences (Dell, 1990). Finally, inflectional affixes are more likely to migrate than are derivational affixes (Garrett, 1982). Collectively, these observations indicate that the constituent structure of morphologically complex words is available to the production mechanism and are consistent with the claim that inflectional and derivational forms may be treated differently (see also Badecker & Caramazza, 1989; Miceli & Caramazza, 1988).

In Experiments 4, 5 and 6, an experimental task inspired by the pattern of speech errors in spontaneous speech was developed in order to provide converging evidence for the claim that the morphological constituents of a word can be available to a processing mechanism. The *segment shifting task* entails deliberately shifting segments from a source word to a target word and rapidly naming the product aloud. The experimental manipulation exploits the fact that the morphemic structure of many words is not wholly transparent and that the same sequence of letters (e.g., ER) can function morphemically in one context (e.g., DRUMMER) and nonmorphemically in another (e.g., SUMMER). Letter sequences which are morphological in the context of some source words and nonmorphological in the context of others were shifted onto the same target word. Pronunciation latencies for the same targets formed from morphemic and nonmorphemic source words are compared. The segment shifting procedure used in Experiment 4 is depicted in Figure 1.

Methods

Subjects. Twenty-six students at the University of Belgrade participated in the experiment in partial fulfillment of the requirements for an Introductory Psychology course. All had experience with reaction time studies but none had participated in previous experiments in this study. The data from nine addition subjects were eliminated because their error rates exceeded 20%.

Stimulus materials. Sixteen pairs of Serbian words were constructed for each of two morphological types and these constituted the source words for Experiment 4. Each pair of source words included a morphologically complex word composed of a base morpheme and a morphological suffix and a morphologically simple control word. The control word ended with the same sequence of letters that functioned morphemically in its pair. Morphemic and nonmorphemic endings were controlled for phonemic and syllabic structure (Tyler & Nagy, 1989). The Serbian analog of inflected words such as WINNING and matched morphologically simple words such as INNING constituted an inflectional type pair. For example, inflectional source words consisted of masculine singular instrumentals such as PRESOM, which means *press*, and nonmorphemic controls consisted of morphologically simple words ending with the same sequence of letters without a morphemic function such as PRELOM, which means *fracture*, in nominative case. Note that the OM sequence appeared on morphemic and nonmorphemic source words of equal length and that source and target words were semantically unrelated.

A second morphological type consisted of homographic morpheme affixes. Pairs of source words consisted of morphologically complex source words with morphological affixes that were compatible with the target word (same syntactic category and gender) and morphologically complex source words that were not. That is, the Serbian analog of nominal or verbal S was shifted to another word of the same (consistent) or a different (inconsistent) word class. For example, the nominative plural I from CEVI, meaning *pipes*, or the third person singular I from CEDI, meaning *he wrings* was shifted to the target word RAD in order to form the word RADI meaning *he works*.

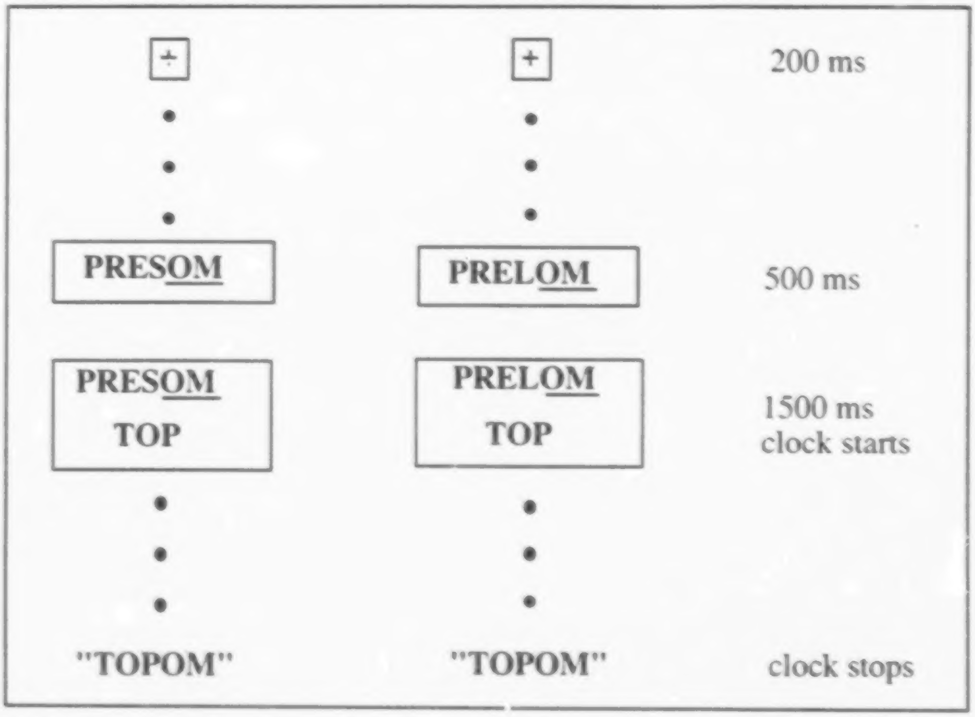


Figure 1. The original segment shifting procedure.

Note that in both source words the I is morphemic. What differs for homographic affixes is the consistency or inconsistency of the syntactic category of the source word and the target word. Source words for the segment shifting experiments are described in Table 1 and are listed in Appendix 1.

Procedure. Following the presentation of a fixation point for 200 ms, a source word with a portion underlined appeared for 500 ms. Immediately afterwards, the target word appeared below the source word and a clock started. Both words remained visible for 1500 ms. A blank field followed the display and lasted for 2000 ms.

Subjects were instructed to segment and shift the designated segment from a source word onto a target word and to name the new result aloud as rapidly as possible. For example, the OM of the source word PRESOM was underlined and subjects were instructed to shift that sequence of letters to the target word TOP in nominative case in order to produce TOPOM, which means *cannon* in instrumental case. Onset to vocalization was measured and errors were recorded. A sequence of 13 practice items preceded the experimental list which included eight tokens each in the morphemic-nonmorphemic and morphemic-incompatible conditions.

Results and Discussion

Means for Experiment 4 are summarized in Table 6. All correct scores less extreme than 3 SD from the mean for each subject were included in an analysis of variance (approximately 14% of all scores were eliminated) and revealed a significant effect of morphological type (inflection/homograph) [$F1(1,25) = 7.53$, $MS_e = 1278$, $p < .01$; $F2(1,30) = 4.85$, $MS_e = 1222$, $p < .04$]. The interaction of morphological status and morphological type was significant in the analysis by subjects [$F1(1,25) = 11.89$ $MS_e = 437$, $p < .003$], but was only marginally significant in the analysis by items [$F2(1,30) = 2.96$, $MS_e = 1550$, $p < .10$]. The effect of morphological status was not significant. A planned comparison between morphological and nonmorphological segments was significant for the inflectional type of affix [$F1(1,25) = 10.44$, $MS_e = 585$, $p < .001$; $F2(1,15) = 3.25$, $MS_e = 1115$, $p < .09$], but not for the homographic type. No effects were significant for errors.

For the homographic morpheme type, where the consistency or inconsistency of the syntactic category of the source word and the target word was varied, no significant effects of consistency (-6 ms) were observed. Shifting rates for "I" segments derived from verbal and nominal source words were statistically equivalent.

The outcome of Experiment 4 was that morphological segments were shifted from source words to targets words more rapidly than their phonologically-matched controls, but that syntactically congruent and incongruent morphological affixes did not differ. This result suggests that the component structure of morphologically complex words is available to the language processing mechanism and that morphemes as contrasted with phonemes are the relevant units.

Table 6. *Segment shifting times and errors for morphological affixes and their nonmorphological or incompatible controls in Experiment 4.*

	SHIFTED SEGMENT		
	MORPHEMIC	NONMORPHEMIC or INCOMPATIBLE	DIFFERENCE
MORPHOLOGICAL TYPE			
INFLECTION	PRESOM	PRELOM	
	582 11%	604 12%	22 1
HOMOGRAPH	CEDI	LEVI*	
	577 13%	571 13%	-6 0

*incompatible morpheme

The observed effect could reflect the lexical representation of morphological structure such as the process of segmenting, from the source, a sequence of letters that constitutes a morphemic component and of affixing that sequence to the target. That is, *segmentation* of morphological units could underlie the effect. Similarly, it is possible that the internal structure of words composed of multiple morphemes differ in their *coherence* relative to morphologically simple words. It should be pointed out that coherence, defined in terms of sequential probabilities between letters, is a not plausible account, because the composition of morphemic and nonmorphemic sequences was well matched in this study (see also Rapp, 1992). Nevertheless, the representation of morphologically complex words may encompass their sub-word units, and morphological coherence may be relevant. In summary, morphemic affixes were more easily segmented from a source word than were nonmorphemic controls presumably because the availability of sublexical morphological components determined morphological coherence. In effect, the imposed shifting of letter sequences from morphologically-simple words is difficult because it is arbitrary, whereas the shifting of letter sequences from morphologically-complex words is relatively easy because it is principled and follows morphological structure.

EXPERIMENT 5

The purpose of Experiment 5 was to replicate the results of the previous experiment and to allow a new comparison between inflectional and derivational morphological types. In an attempt to increase the magnitude of the effects observed in the previous experiments, the segment of the source word that subjects had to shift was not specified when the source word appeared. Instead, it was indicated 750 ms later and was simultaneous with the appearance of the target word. The comparison between inflectional and derivational affixes was again examined. If the constituent structure of inflections is more transparent than that of derivations, then effects should be more systematic for inflections. Finally, if the segment shifting effect is sensitive to strategies imposed by the subject and if subjects anticipate segmenting morphological affixes, then limiting preparation time before the onset of the target may increase the magnitude of the effect because the component structure of the morphemic source word, but not its control, will be available before it is visually specified.

Methods

Subjects. Twenty-four students at the University of Belgrade participated in Experiment 5 in partial fulfillment of the requirements for an Introductory Psychology course. All had experience with reaction time studies but none had participated in previous experiments in this study. No subject's data were eliminated because of error rates in excess of 20%.

Stimulus materials. Eighteen pairs of Serbian words were constructed for each of two morphological types and these constituted the source words. Each pair of source words included a morphologically complex word (composed of a base morpheme and a morphological suffix) and a morphologically simple control word. The control word ended with the same sequence of letters that functioned morphemically in its pair. Inflectional type source words consisted of first person singular verbs ending in EM such as KRADEM, which means *I steal,* and nonmorphemic controls consisted of morphologically simple words ending with the same sequence of letters without a morphemic function such as BADEM, which means *almond,* in the nominative case. Note that the EM sequence appeared on morphemic and nonmorphemic source words whose length differed by no more than one letter.

A second morphological type consisted of agentives which are derivational morphemes. These pairs of source words consisted of morphologically complex source words and morphologically simple source words ending in the sequence AR or AC. For example, derivational source words consisted of agents such as CUVAR, meaning *guard,* in nominative case and nonmorphological controls consisted of morphologically simple words such as STVAR, meaning *thing,* in nominative. In both cases, the AR was shifted to the target word RAD in order to form the word RADAR, meaning *worker.* Subjects were instructed to add the shifted segment from the source word to the target word and to name it aloud.

Procedure. In an attempt to increase the size of the effect observed in the previous experiment, the presentation conditions of Experiment 5 were modified. The segment of the source word that subjects had to shift was not indicated at the same time that the source word appeared. That is, the source word first appeared alone and without underlining. After 750 ms, the target word appeared below the source word, the segment of the source word that subjects had to shift was underlined, and a clock started. A blank field followed the display and lasted for 2000 ms.

Subjects were instructed to segment and shift the designated segment from a source word onto a target word and to name the new result aloud as rapidly as possible. For example, the EM of the source word JEDEM was underlined and subjects were instructed to shift that sequence of letters to the target word KUJE in order to produce KUJEM, which means *I hammer*. Onset to vocalization was measured and errors were recorded. A sequence of 13 practice items preceded the experimental list which included nine tokens of morphemic and nonmorphemic source words in the inflectional and derivational conditions.

Results and Discussion

An analysis of variance on correct latencies less extreme than 3 SD from the mean for each subject (approximately 6% of all responses were eliminated) revealed significant effects of morphological type (inflection/derivation) [$F1(1,23) = 13.78$, $MS_e = 2487$, $p < .002$] and morphological status (morpheme/nonmorpheme) [$F1(1,23) = 9.1$, $MS_e = 913$, $p < .007$] by the subjects' analysis but only morphological type approached significance by the items' analysis [$F2(1,34) = 2.75$, $MS_e = 5139$, $p < .11$]. With the error measure, neither the main effect nor the interaction of affix by type approached significance. Means are summarized in Table 7.

Numerical differences for agentive derivational affixes were reduced and in the opposite direction relative to those of inflectional affixes although there was no significant interaction. Nevertheless, a planned comparison conducted on means for each subject indicated that inflectional affixes were shifted faster than their nonmorphological controls [$F1(1,23) = 8.15$, $MS_e = 1201$, $p < .009$] and a test conducted on means for each item showed the same trend [$F2(1,17) = 2.68$, $MS_e = 2393$, $p < .12$]. No effects were significant for derivational affixes, however.

The outcome of Experiment 5 was that inflectional but not derivational segments were shifted from source words to targets words more rapidly than from their phonologically-matched controls. This result suggests that the component structure of morphologically complex words is sometimes available to the language processing mechanism and again, that base morphemes as contrasted with phonemes are the relevant units. Numerically, the effect was comparable to that of Experiment 4 suggesting that restricted preparation time did not alter the processes involved in this task. Times were longer overall but error rates decreased. Importantly, the relation between speed and accuracy across experiment did not differ by experimental condition.

The linguistic productivity and lexical structure of inflectional as contrasted with derivational formations noted above leads one to expect inflectional affixes to show enhanced effects relative to derivational affixes and the effect of morphological status was significant only for inflections in Experiment 5. The sixth experiment in this series also compares inflections and derivations in a more complex version of the segment shifting task.

Table 7. *Segment shifting times and errors for morphological affixes and their nonmorphological controls in Experiment 5.*

		SHIFTED SEGMENT		
		MORPHEMIC	NONMORPHEMIC	DIFFERENCE
MORPHOLOGICAL TYPE				
	INFLECTION	JEDEM	BEDEM	
		781	809	28
		5%	4%	-1
	DERIVATION	ZIDAR	KADAR	
		753	761	8
		4%	5%	1

EXPERIMENT 6

An attempt at replication of differences in inflectional and derivational processing with different materials necessitated a modification of the segment shifting procedure described above. In this experiment, as in the previous segment shifting experiments, subjects had to shift the affix from the source word to the target word. In contrast to the procedure of the previous experiments, in Experiment 6, subjects had to delete the original affix on the target word before substituting the shifted segment. As in Experiments 4 and 5, subjects had to name the resulting word aloud. The addition of this step whereby subjects had to delete the original affix (or its orthographically and phonemically matched control) rendered the task more difficult but it permitted the comparison of morphological constructions for inflectional and derivational formations to be expanded.

Methods

Subjects. Twenty-six students from the same population as those in previous experiments participated in Experiment 6. None had participated in previous experiments in this study.

Stimulus materials. Materials consisted of thirty-six Serbian word pairs including equal numbers of inflectional and derivational morphological types and their nonmorphemic controls. As in the previous experiment, the inflectional type consisted of first person singular verbs such as PROGONIM, meaning *I capture*, and their nonmorphemic controls such as SINONIM, meaning *synonym*. They were shifted to inflected targets such as DELE, meaning *they share*. In the present experiment, in order to respond DELIM, meaning *I share*, subjects had to delete the original affix (viz., E) and substitute the IM affix. The derivational type contrast consisted of singular diminutives ending in ICA such as BASTICA, meaning *little garden*, and their controls such as KOSTICA, meaning *seed*. They were shifted to target such as BUBA, meaning *bug* and subjects responded BUBICA, meaning *little bug*.

Procedure. The procedure of Experiment 6 was like that of Experiment 5 (but not 4), in that the segment of the source word that subjects had to shift was not specified at the same time that the source word appeared. Instead, it was indicated after 750 ms when the target word appeared. However, in both the inflectional and derivational conditions of Experiment 6, subjects had to drop the original (morphemic) affix on the target and to substitute the affix from the source word. That is, the final vowel on words such as BACE and BUBA was deleted before adding IM or ICA respectively. Finally, filler trials in which no portion was underlined were also included. In those cases, subjects were required to repeat the target word in its original form.

Nine tokens in the morphemic and nonmorphemic conditions were included for both the inflectional and derivational conditions.

Results and Discussion

An analysis of variance on correct latencies less extreme than 3 SD from the mean (so that approximately 6% of all responses were eliminated) revealed a significant effect of morphological type (inflection/derivation) [$F1(1,25) = 60.03$, $MS_e = 3295$, $p < .001$; $F2(1,34) = 17.12$, $MS_e = 8177$ $p < .001$] and a marginally significant interaction of morphological status and morphological type [$F1(1,25) = 10.02$, $MS_e = 2960$, $p < .005$; $F2(1,34) = 2.77$, $MS_e = 7182$, $p < .10$]. The effect of morphological status just missed significance with subjects as a random variable [$F1(1,25) = 4.12$, $MS_e = 3250$, $p < .053$]. Means are summarized in Table 8.

Table 8. *Segment shifting times and errors for morphological affixes and their nonmorphological controls in Experiment 6.*

	SHIFTED SEGMENT		
	MORPHEMIC	NONMORPHEMIC	DIFFERENCE
MORPHOLOGICAL TYPE			
INFLECTION	BAJAM	SAJAM	
	829	886	57
	9%	6%	-3
DERIVATION	BAŠTICA	KOSŤICA	
	776	765	-9
	3%	3%	0

The effect of morphological status of affix (56 ms) was significant for inflectional type pairs [$F1(1,25) = 14.01$, $MS_e = 2960$, $p < .001$)] by subjects and was marginally significant by items [$F2(1,34) = 3.24$, $MS_e = 7182$, $p < .08$)]. For derivational pairs, the effect was in the opposite direction (-11 ms) and was not significant [$F1(1,25) = .54$]. The significant effect for inflections and the nonsignificant change in direction for derivations produced the marginally significant interaction of affix by morphological type. For errors, the effect of morphological type was significant by subjects [$F1(1,25) = 9.82$, $MS_e = 575$, $p < .005$; $F2(1,34) = 7.73$, $MS_e = 398$, $p < .09$] but the main effect of morphological status and the interaction of contrast by type did not approach significance. Because latency and error patterns for the targets following inflected primes in Experiment 6 suggested a speed accuracy tradeoff, correlations between measures were computed. Neither the correlations for morphemic and nonmorphemic conditions separately nor the pooled correlation approached significance. Evidently, latencies did not decrease as errors increased.

The results of Experiment 6 are consistent with the segment shifting results of the previous experiments whereby morphological segments are shifted faster than their nonmorphemic controls. The pattern of errors goes in the opposite direction but it was not statistically significant nor was it produced by a speed-accuracy tradeoff. Although the results with items as a random factor are weak, the pattern was replicated with (a) the inflectional affixes for instrumental nouns in Experiment 4 (b) first person singular verbs in Experiments 5 and 6. The set of experimental materials for Experiment 6 required a modification to the experimental procedure whereby the original affix on the target word had to be deleted before the shifted segment could be appended and it allowed a valuable replication of the previous results. Specifically, the effect of segment shifting was significant for inflectional pairs but not for derivational pairs. These results are consistent with the linguistic distinction between morphological types noted above and with the pattern of production error whereby inflections enter into speech errors more frequently than do derivations (Garrett, 1980). This finding suggests that the morphological structure of inflectional and derivational formations does differ.

GENERAL DISCUSSION

In the repetition priming paradigm, the pattern of facilitation among lexical decision latencies for target words that were preceded by morphological relatives provided evidence that skilled readers of Serbian are sensitive to the constituent structure of morphologically complex words. It was not necessary for identical forms to be repeated in order to reduce target decision latencies. Repetition of the same base morpheme in a different but related morphologically-complex word also produced facilitation. Evidence of morphological relatedness in repetition priming is consistent with the results of similar studies conducted across a variety of languages and morphological contexts, and generally, it is interpreted as evidence that morphology is represented in the lexicon.

Similarly, the failure to find facilitation in lexical decision among target pseudowords that were preceded by other pseudowords constructed from the same meaningless base morpheme and real morphemic suffixes, or by words constructed from the same meaningful base morpheme in an illegal combination with a real affix, is consistent with the outcome of other studies that have used this experimental task. Although small facilitation effects for pseudoword targets sometimes have been reported in lexical decision with repetition priming (e.g., Bentin & Feldman, 1990), it is never the case that pseudoword effects are numerically larger than word effects and most typically they are smaller. It has been suggested that for pseudowords, under some encoding conditions, the advantage of repeating the same or a very similar orthographic and phonemic pattern is offset by the inappropriateness of responding "word" to a familiar pseudoword pattern (Balota & Chumbley, 1984; Duchek & Neely, 1989; Feustel, Shriffrin & Salasoo, 1983). That is, familiarity offsets any advantage associated with repeating a "no" response.

The present experiments conducted with Serbian materials permitted a rigorous comparison of two types of morphological formations. When inflectionally- and derivationally-related prime-target pairs were compared, significantly greater facilitation was observed for inflectional relatives than for derivational relatives. This finding was observed in Experiment 1, in which derivational formations differed in word class from inflectional formations but were equally similar with respect to phonological and orthographic overlap, in Experiment 2, in which all formatives were verbs and targets following derivationally-related primes differed with respect to the addition of either a prefix or a suffix, and in Experiment 3, in which one half of the primes were perfectly matched for overlap as well as word class with their targets and one half shared one letter more in the inflectional condition than in the derivational condition.

When derivationally-related prime-target pairs were compared with first presentation, significant facilitation was observed for agentives in Experiment 1, for prefixed targets in Experiment 2, and for mismatched pairs in Experiment 3. Facilitation following derivational primes was not significant in the analysis by subjects for infixed items (Experiment 2). The planned comparison by items for matched items (Experiment 3) was not significant although the latency differences were consistent with a pattern of facilitation. The reliability of facilitation from derivationally-related primes may be low and the no prime baseline may overestimate the magnitude of facilitation (but see discussion of results for Experiment 1-3). However, the same pattern was observed in three experiments. Moreover, when the planned comparisons for no prime and derivational conditions in the three experiments (five conditions) were combined into one statistical test (Winer, 1971 p.49), results indicated that facilitation was significant [$\chi^2(10) = 48.21$ $p < .001$ for subjects and $\chi^2(10) = 38.97$] $p < .001$ for items. In summary, although there is a tendency towards facilitation of targets following derivational primes, because targets always occurred slightly later in the session than their primes, the facilitation effects with derivational relatives should be interpreted with caution.

It is unlikely that the effect of morphological relatedness can be described in terms of the pattern of activation between the orthographic and perhaps phonological units that constitute a word. This claim is based on a) the pattern of facilitation for morphologically-related prime-target pairs in which the base morpheme does not always retain the same form, b) the absence of facilitation among morphologically-unrelated prime-target pairs that are similar in form, c) statistically significant differences in patterns of facilitation to targets primed by inflectional and derivational relatives that are matched or nearly matched to the target for similarity of form, and d) the effect of morphological status on segment shifting for inflectional affixes but not for derivational affixes. The separate bases for these claims will now be summarized.

In previous repetition priming studies, changes in spelling or pronunciation between morphologically-related prime and target did not diminish the magnitude of facilitation to target decision latencies relative to morphologically-related prime-target pairs with no change. For example, Serbian forms that undergo palatalization (e.g., NOZI), forms with letter deletion (e.g., PETKU) and regular forms (e.g., NOGOM) all produced equivalent target (e.g., NOGA, PETAK) facilitation (Feldman & Fowler, 1987). Similar results in repetition priming have been reported in English for moderately irregular forms such as SLEPT- SLEEP (Fowler et al., 1985; Napps, 1989; Stolz & Feldman, in press; cf. Stanners et al., 1979). In addition, recognition latency to inflected verb forms was sensitive to frequency of citation forms (and cumulative frequency for all regular forms) both when they differed in spelling (Kelliher & Henderson, 1990) and when spelling was preserved (Katz, Rexer & Lukatela, 1991; Nagy, Anderson, Schommer, Scott & Stallman, 1989). That is, a contribution of both citation frequency and cumulative frequency of morphologically-related forms to recognition latency was observed even when the orthographic and phonological form of the base morpheme was not preserved. Equivalent patterns of influence for morphologically-related words with differing orthographic and phonological form and for words with similar form are problematic for any model that assumes that the base morpheme alone or a pattern of activation over its letter or phoneme units underlies morphological effects. In summary, patterns of facilitation in repetition priming suggest that the underlying morphemic representation is abstract enough to tolerate at least moderate orthographic and phonological variation.

Whereas formal similarity of morphologically unrelated words can produce inhibition in some presentation formats when items are presented close in succession (Grainger, Colé, & Segui, 1990; Laudanna et al., 1992; Segui & Grainger, 1990; Stolz & Feldman, in press), at long lags it is the case that the formal similarity of morphologically unrelated primes and targets (e.g., pairs such as DIET and DIE) does not result in priming either in English (Hanson & Wilkenfeld, 1985; Napps & Fowler, 1987; Stolz & Feldman, in press) or in Serbian (Feldman & Andjelković, 1992; Feldman & Moskovljević, 1987;). For example, for prime-target pairs formed from unrelated homographic base morphemes (e.g., BOR) such as BORAMA (dative plural of BORA meaning *wrinkle*) and BOROVI (nominative plural of BOR meaning *pine*), no effect of formal similarity was observed for inflectionally complex words at lags of ten items in repetition priming (Feldman & Andjelković, 1992). Stated generally, whereas orthographically similar form plays a role at short lags, at long lags, as in the repetition priming task, there was neither a facilitative nor inhibitory effect on the target of orthographic and phonological similarity between morphologically-unrelated prime and target.

The primary finding of Experiments 1, 2 and 3 was that inflectional primes produced significantly greater facilitation than did derivational primes and that derivational primes produced facilitation relative to the no prime condition. This outcome was observed under experimental conditions that a) perfectly matched formal overlap of prime and target but left word class free to vary (Exp. 1), b) perfectly controlled word class by using only verb forms and manipulated position in which affix was added and c) perfectly controlled word class by using only verb forms but matched letter overlap of inflectional and derivational relatives on only one half of the items (Exp. 3).

It is evident that the lexical representation of inflectional and derivational formations must differ. Several accounts have been proposed. It has been suggested that, in the lexicon, the linkage between whole word forms that share a base morpheme is stronger for inflectionally-related forms than for derivationally-related forms or that the connections between components must be stronger for inflectional than for derivational formations. This is consistent with the linguistic claim that the component structure of inflectionally-related forms is more transparent than that of derivationally-related forms. Alternatively, as noted above, inflectional formations tend to share a base morpheme and stem and differ with respect to inflectional affix whereas derivational formations tend to share a base morpheme and differ with respect to derivational affixes and stem. Accordingly, if both stems and bases are taken as units to be activated in repetition priming, then the difference between inflections and derivations could reflect redundant activation for inflections relative to derivations. Results of Experiment 3 indicate that this account is incomplete, at best. Inflectional relatives like NOSIM and derivational relatives like NOSAM share both base morpheme and stem but they did not produce equivalent target facilitation.

Results of the segment shifting task also provide compelling evidence that the morphological status of a word's constituents influences performance in recognition tasks. Although differences between inflectional and derivational affixes were weaker in this task, in that analyses by items tended to have significance levels around $p < .10$, a similar outcome was observed in three experiments with different manipulations on inflectional affixes. Importantly, the results were consistent with those obtained in repetition priming. In the segment shifting task, morphological effects were more reliable for inflectional affixes relative to their controls than for derivational affixes and their controls. Phonological and orthographic properties of a source word were matched in the experimental design and sequences that created morphological segments were manipulated more efficiently than nonmorphological controls over a variety of inflectional environments. The morphological origins of a segment were evident despite the fact that all of the responses articulated by subjects were frequent words and that responses were the same in the morphological and nonmorphological conditions.

In that the internal structure of words formed with inflectional affixation may be more transparent than that of words with derivational affixation, and effects were more reliable for inflections than for derivations, an account of the segment shifting experiments based on ease of segmentation is plausible. Similarly, the emphasis could be on coherence at the boundary between sublexical word units. In languages such as Serbian (and English) in which morphemes are concatenated to form complex forms, there is a temptation to describe morphemes in terms of boundaries between orthographic or phonological *sequences* of units. As should be obvious from the present discussion, however, these lexical representations must be sufficiently abstract to accommodate changes in form as well as the word context of which the sequence is a part. Admittedly, it is difficult to distinguish between segmentation of and coherence between morphemic components when morphology is concatenated.

An investigation of morphological effects in a nonconcatenative language such as Hebrew is less amenable to an account based on sequence, however. In Hebrew, the root or base morpheme of a word is represented by a discontinuous pattern of (usually) three consonants. Vowels carrying an inflectional function are infixed between these units. Consequently the base morpheme is not realized as an uninterrupted unit within the word. It is necessarily abstract with respect to phonological (and sometimes orthographic) patterning and yet, effects of morphological relatedness have been demonstrated both in repetition priming (Bentin & Feldman, 1990; Feldman & Bentin, 1994) and in segment shifting (Feldman, Frost & Pnini, in press). In Hebrew, ease of segmentation rather than coherence may provide the more accurate account.

Alternatively, the locus of the segment shifting effect could reflect strategies that vary on a trial by trial basis and reflect compatibility between source and target word. Accordingly, word class compatibility effects for affixes from source and target words would be anticipated. The CEVI-CEDI comparison in Experiment 4 indicated that when affixes were shifted from a source word of the same or a different word class as the target, latencies were equivalent. Similarly, the results of segment shifting experiments in Hebrew (Feldman, Frost & Pnini, in press) indicate that morphological effects can be observed when the affix is shifted to a meaningless target string. Segment shifting effects are not expected on pseudoword targets if the effect reflects compatibility. Evidently, the locus of the segment shifting effect is not yet well understood but, at this point, a lexical locus tied to morphemic as distinguished from orthographic components seems likely.

In summary, two very different experimental paradigms provide strong support for the psychological processing of the morpheme and for a distinction between the processing of inflectional and derivational formations. Similarity of form defined by orthographic and phonological overlap of morphologically-related primes and targets is not a necessary condition to produce facilitation. Similarity of form in the absence of morphological relatedness is not a sufficient condition to produce target facilitation or inhibition at long lags. Patterns of activation over orthographic or phonological units cannot describe morphological effects. Importantly, they cannot account for the differences between inflections and derivations when semantic similarity is controlled. Evidently processing of a word is sensitive to that word's constituent morphemic structure.

REFERENCES

Anderson, S. R. (1982). Where's morphology? *Linguistic Inquiry, 13,* 571-612.

Aronoff, M. (1976). *Word formation in generative Grammar.* Cambridge, MA: MIT Press.

Balota, D. A., & Chumbley, J. I. (1984). Are lexical decision a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception and Performance, 10,* 340-357.

Badecker, W., & Caramazza, A. (1989). A lexical distinction between inflection and derivation. *Linguistic Inquiry, 20,* 108-116.

Bentin, S., & Feldman, L. B. (1990). The contribution of morphological and semantic relatedness to repetition priming at short and long lags: Evidence from Hebrew. *Quarterly Journal of Experimental Psychology, 42A,* 693-711.

Bergman, B., Hudson, P., & Eling, P., (1988). How simple complex words can be: Morphological processing and word representations. *Quarterly Journal of Experimental Psychology, 40A,* 41-72.

Burani, C., & Laudanna, A., (1992). Units of representation for derived words in the lexicon. In R. Frost and L. Katz (Eds.) *Orthography, phonology, morphology and meaning* (pp. 361-376). Amsterdam: Elsevier Science Publishers B.V.

Butterworth, B. (1983). Lexical representation. In B. Butterworth (Ed.), *Language production* (Vol. 1). London and San Diego: Academic Press.

Bybee, J. H. (1985). *Morphology.* Amsterdam: John Benjamins.

Caramazza, A., Laudanna, A., & Romani, C. (1988). Lexical access and inflectional morphology. *Cognition, 28,* 287-332.

Caramazza, A., Miceli, G., Silveri, M. C., & Laudanna, A. (1985). Reading mechanisms and the organization of the lexicon: Evidence from acquired dyslexia. *Cognitive Neuropsychology, 2,* 81-114.

Cutler, A. (1980). Errors of stress and intonation. In V. A. Fromkin (Ed.), *Errors in linguistic performance.* New York: Academic Press.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review, 93,* 285-321.

Dell, G. S. (1990). Effects of frequency and vocabulary type on phonological speech errors. *Language and Cognitive Processes, 5,* 313-349.

Duchek, J. N., & Neely, J. H. (1989). A dissociative word-frequency × levels-of -processing interaction in episodic recognition and lexical decision tasks. *Memory & Cognition, 17,* 148-162.

Emmorey, K. D. (1985). Auditory morphological priming in the lexicon. *Language and Cognitive Processes, 4,* 73-92.

Feldman, L. B. (1992). Morphological relationships revealed through the repetition priming task. In M. Noonan, P. Downing, & S. Lima (Eds.), *Literacy & Linguistics.* Amsterdam/Philadelphia: John Benjamins Pub. Co.

Feldman, L. B., & Andjelković, D. (1992). Morphological analysis in word recognition. In R. Frost & L. Katz (Eds.), *Phonology, orthography, morphology and meaning* (pp. 343-360). Amsterdam: Elsevier Science Publishers B.V.

Feldman, L. B., & Bentin, S. (1994). Morphological analysis of disrupted morphemes: Evidence from Hebrew. *Quarterly Journal of Experimental Psychology, 47A,* 407-435.

Feldman, L. B., Frost, R. & Pnini, D. (in press). Morphemes need not be contiguous units. *Journal of Experimental Psychology: Learning, Memory and Cognition.*

Feldman, L. B., & Fowler, C. A. (1987). The inflected noun system in Serbo-Croatian: Lexical representation of morphological structure. *Memory & Cognition, 15,* 1-12.

Feldman, L. B. & Fowler, C. A. (1987b) Morphemic segments shift faster than their nonmorphemic controls. Paper presented to the Psychonomic Society, Seattle, WA.

Feldman, L. B., & Moskovljević, J. (1987). Repetition priming is not purely episodic in origin. *Journal of Experimental Psychology: Learning, Memory & Cognition, 13,* 573-581.

Feustel, T. C., Shriffrin, R. M., & Salasoo, A. (1983). Episodic and lexical contributions to the repetition priming effect in word identification. *Journal of Experimental Psychology: General, 112,* 309-346.

Fowler, C. A., Napps, S. E., & Feldman, L. B. (1985). Relations among regular and irregular morphologically related words in the lexicon as revealed by repetition priming. *Memory & Cognition, 13,* 241-255.

Fromkin, V. A. (1973). *Speech errors as linguistic evidence.* Mouton: The Hague.

Garrett, M. F. (1976). Syntactic processes in language production. In R. J. Wales & E. C. T. Walker (Eds.), *New approaches to language mechanisms* (pp. 231-256). Amsterdam: North Holland.

Garrett, M. F. (1980). A perspective on research in language production. In E. C. Mehler, T. Walker, & M. F. Garrett (Eds.), *Perspectives on mental representation* (pp. 179-220). Hillsdale, NJ: Lawrence Erlbaum Associates.

Garrett, M. F. (1982). Production of speech. Observations from normal and pathological language use. In A. Ellis (Ed.), *Normality and pathology in cognitive functions* (pp. 19-76). London: Academic Press.

Grainger, J., Colé P., & Segui, J. (1991). Masked morphological priming in visual word recognition. *Journal of Memory & Language, 30*, 370-384.

Hanson, V. L., & Feldman, L. B. (1989). Language specificity in lexical organization: Evidence from deaf signers' lexical organization of ASL and English. *Memory & Cognition, 17*, 292-301.

Hanson, V. L., & Wilkenfeld, D. (1985). Morphophonology and lexical organization in deaf readers. *Language and Speech , 28*, 269-280.

Henderson, L. (1985). Toward a psychology of morphemes. In A. W. Ellis (Ed.), *Progress in the Psychology of Language* (pp. 15-72). London: Erlbaum.

Hudson, P. T. W. (1990). What's in a word? Levels of representation and word recognition. In D. A. Balota, G. B. Flores d'Arcais & K. Rayner (Eds.), *Comprehension processes in reading*. Hillsdale: Erlbaum.

Katz, L., Rexner, K., & Lukatela, G. (1991). The processing of inflected words. *Psychological Research, 53*, 25-31.

Kelliher, S., & Henderson, L. (1990). Morphologically based frequency effects in the recognition of irregularly inflected verbs. *British Journal of Psychology, 81*, 527-539.

Kirsner, K., Milech, D., & Standen, P. (1983). Common and modality-specific processes in the mental lexical. *Memory & Cognition, 11*, 621-630.

Laudanna, A., Badecker, W., & Caramazza, A. (1992). Processing inflectional and derivational morphology. *Journal of Memory & Language, 31*, 333-348.

Lyons, J. (1977). *Semantics* (Vol. 2). Cambridge: Cambridge University Press.

Miceli , G. , & Caramazza , A. (1988). Dissociation of inflectional and derivational morphology. *Brain and Language, 35*, 24-65.

Nagy, W., Anderson, R. C., Schommer, M., Scott, J. A., & Stallman, A. C. (1989). Morphological families in the internal lexicon. *Reading Research Quarterly, 24*, 263-282.

Napps, S. E. (1989). Morphemic relationships in the lexicon: Are they distinct from semantic and formal relationships? *Memory & Cognition, 17*, 729-739.

Napps, S. E., & Fowler, C. A. (1987). Formal relationships among words and the organization of the mental lexicon. *Journal of Psycholinguistic Research, 16*, 257-272.

Neely, J. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. In D. Besner & G. W. Humphreys (Eds.), *Basic processes in reading: Visual word recognition*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Partridge, M. (1964). Serbo-Croatian practical grammar and reader. Belgrade: University of Belgrade Press.

Radeau, M. (1983). Semantic priming between spoken words in adults and children. *Canadian Journal of Psychology, 37*, 547-556.

Rapp, B. C. (1992). The nature of sublexical orthographic organization: The bigram trough hypothesis examined. *Journal of Memory and Language, 31*, 33-53.

Scalise, S. (1984). *Generative morphology*, Dordrecht: Foris Publications.

Schriefers, H., Friederici , A., & Graetz, P., (1992). Inflectional and derivational morphology in the mental lexicon: Symmetries and asymmetries in repetition priming. *Quarterly Journal of Experimental Psychology, 44*, 373-390.

Segui, J., & Grainger, J. (1990). Priming word recognition with orthographic neighbors: Effects of relative prime-target frequency. Journal of Experimental Psychology: Human Perception and Performance, 163 65-76.

Slowiaczek, L. M. (1994). Semantic priming in a single-word shadowing task. *American Journal of Psychology, 107*, 245-260.

Stanners, R. F., Neiser, J. J., Hernon, W. P., & Hall, R. (1979). Memory representation for morphologically related words. *Journal of Verbal Learning and Verbal Behavior, 18*, 399-412.

Stemberger, J. P. (1984). Structural errors in normal and agrammatic speech. *Cognitive Neuropsychology, 1*, 281-313.

Stemberger, J. P. (1985). An interactive activation model of language production. *Progress in the psychology of language* (pp. 143-186). London: Lawrence Erlbaum Associates.

Stolz, J. A., & Feldman, L. B. (in press). The role of orthographic and semantic transparency of the base morpheme in morphological processing. In L. B. Feldman (Ed.), *Morphological aspects of language processing*. Hillsdale: Lawrence Erlbaum Associates.

Tyler, A., & Nagy, W. (1989). The acquisition of Engl[illegible] derivational morphology. *Journal of Memory & Languag[illegible]*, [illegible], 649-667.

Winer, B. J. (1971). *Statistical principles in experimental design*. New York: McGraw-Hill.

FOOTNOTES

[*] Appears in *Journal of Memory and Language, 33*, 442-470 (1994).

[†] Also State University of New York at Albany.

BEST COPY AVAILABLE

APPENDIX 1

Experiment 1 materials

TARGET	IDENTITY	INFLECTION	DERIVATION
blud	blud	bludom	bludim
bol	bol	bolom	bolan
branim	branim	braniš	branik
brod	brod	brodu	brodi
broj	broj	broju	broji
bura	bura	burom	buran
čud	čud	čudom	čudim
cvet	cvet	cvetom	cvetan
deo	deo	delom	delim
doček	doček	dočekom	dočekan
govor	govor	govorom	govorim
grad	grad	gradom	gradim
hlad	hlad	hladom	hladan
igraš	igraš	igram	igrač
lom	lom	lomom	lomim
nosam	nosam	nosaš	nosaš
obaram	obaram	obaraš	obarač
peva	peva	pevam	pevač
plivam	plivam	plivaš	plivač
računa	računa	računam	računar
rad	rad	radom	radim
slikaš	slikaš	slikam	slikar
spava	spava	spavaš	spavač
tragaš	tragaš	tragam	tragač
vaja	vaja	vajam	vajar
vodiš	vodiš	vodim	vodič
zvoni	zvoni	zvonim	zvonik

Experiment 2 materials

	TARGET	IDENTITY	INFLECTION	DERIVATION
PREFIXED:				
	obare	obare	obarim	barim
	ocede	ocede	ocedim	cedim
	očiste	očiste	očistim	čistim
	obodre	obodre	obodrim	bodrim
	oderu	oderu	oderem	derem
	oljušte	oljušte	oljuštim	ljuštim
	iseku	iseku	isečem	sečem
	iskoče	iskoče	iskočim	skočim
	isele	isele	iselim	selim
	ulepe	ulepe	ulepim	lepim
	urade	urade	uradim	radim
	ubodu	ubodu	ubodem	bodem
	zgrabe	zgrabe	zgrabim	grabim
	zbroje	zbroje	zbrojim	brojim
	zbace	zbace	zbacim	bacim
	zdrobe	zdrobe	zdrobim	drobim
	zgnjeće	zgnjeće	zgnjećim	gnjećim
	zbrišu	zbrišu	zbrišem	brišem
	slete	slete	sletim	letim
	smute	smute	smutim	mutim
	sprže	sprže	spržim	pržim
	smrve	smrve	smrvim	mrvim
	stresu	stresu	stresem	tresem
	slome	slome	slomim	lomim
INFIXED:				
	birkaju	birkaju	birkam	biram
	čarnu	čarnu	čarnem	čaram
	dirnu	dirnu	dirnem	diram
	džarnu	džarnu	džarnem	džaram
	gurnu	gurnu	gurnem	guram
	javnu	javnu	javnem	javim
	jurnu	jurnu	jurnem	jurim
	kidnu	kidnu	kidnem	kidam
	kucnu	kucnu	kucnem	kucam
	mere	mere	merkam	merim
	mrdnu	mrdnu	mrdnem	mrdam
	njuše	njuše	njuškam	njušim
	padnu	padnu	padnem	padam
	pirkaju	pirkaju	pirka	piri
	sednu	sednu	sednem	sedam
	sevnu	sevnu	sevnem	sevam
	šušte	šušte	šuškam	šuštim
	svirkaju	svirkaju	svirkam	sviram
	turnu	turnu	turnem	turam
	virkaju	virkaju	virkam	viram
	virnu	virnu	virnem	virim
	vrdnu	vrdnu	vrdnem	vrdam
	živnu	živnu	živnem	živim
	zovnu	zovnu	zovnem	zovem

Expariment 3 materials

TARGET	IDENTITY	INFLECTION	DERIVATION
MATCHED:			
bace	bace	bacim	bacam
ciče	ciče	cičim	cičem
dobiju	dobiju	dobijam	dobijem
klize	klize	klizim	klizam
lupe	lupe	lupim	lupam
nose	nose	nosim	nosam
odbiju	odbiju	odbijam	odbijem
opuste	opuste	opustim	opuštam
sede	sede	sedim	sedam
speru	speru	speram	sperem
vode	vode	vodim	vodam
voze	voze	vozim	vozam
čuče	čuče	čučim	čučnem*
MISMATCHED:			
dirnemo	dirnemo	dirnem	diram
duvnemo	duvnemo	duvnem	duvam
napijamo	napijamo	napijam	napijem
natapamo	natapamo	natapam	natopim
nazivamo	nazivamo	nazivam	nazovem
naturamo	naturamo	naturam	naturnem
obaramo	obaramo	obaram	oborim
obijamo	obijamo	obijam	obijem
odvajamo	odvajamo	odvajam	odvojim
potapamo	potapamo	potapam	potopim
povijamo	povijamo	povijam	povijem
skidamo	skidamo	skidam	skinem
zovnemo	zovnemo	zovnem	zovem

*not included in the analysis

Experiment 4 materials

	MORPHOLOGICALLY COMPLEX WORD	MORPHOLOGICALLY SIMPLE WORD	TARGET
INFLECTIONS			
	slavom	slalom	rad
	presom	prelom	top
	prozom	prolom	grof
	akcijom	aksiom	dan
	maljem	melem	kraj
	bojem	boem	muž
	palcem	parfem	broj
	hicem	harem	nož
	prolećem	problem	čaj
	tučem	totem	konj
	adresom	agronom	kum
	princem	prijem	koš
	arkadom	astronom	lav
	etikom	ekonom	zid
	kafanom	karcinom	zet
	antenom	anatom	sat
HOMOGRAPHIC MORPHEME			
	cedi	cevi	jad
	davi	dani	ćud
	deli	dedi	gost
	gadja	gara	rak
	kida	kipa	lan
	mrzi	mravi	gled
	čami	čari	reč
	košta	koša	nos
	buja	buta	zec
	crpi	crvi	čadj
	krsti	kosti	zub
	prska	prsta	luk
	pada	poda	klub
	mašta	mosta	dom
	prašta	plasta	sir
	gleda	gliba	sin

Experiment 5 materials

	MORPHOLOGICALLY COMPLEX WORD	MORPHOLOGICALLY SIMPLE WORD	TARGET
DERIVATIONS			
	berač	kolač	let
	birač	vrač	ples
	čuvar	stvar	red
	kuvar	ajvar	mlin
	limar	ormar	lug
	merač	žarač	kov
	orač	trač	voz
	pekar	bakar	kalem
	ribar	dabar	stan
	rudar	sudar	dom
	slikar	plakar	grob
	šumar	šamar	brod
	svirač	otirač	glas
	trubač	korbač	kroj
	vidar	radar	drug
	vinar	bunar	zub
	zidar	kadar	lek
	zlatar	litar	put
INFLECTIONS			
	bajam	sajam	kida
	biram	jaram	pita
	derem	bagrem	bere
	diktiram	dijagram	diagram
	jedem	bedem	kluje
	kradem	badem	kaže
	majam	zajam	dira
	orem	harem	digne
	pajam	pojam	buja
	perem	melem	pase
	pijem	prijem	dode
	pletem	sistem	bakče
	postim	kostim	davi
	progonim	sinonim	deli
	tonem	fonem	pišu
	udaram	epigram	kupa
	vajam	najam	pada
	zidam	islam	peva

Experiment 6 materials

	MORPHOLOGICALLY COMPLEX WORD	MORPHOLOGICALLY SIMPLE WORD	TARGET
DERIVATIONS			
	vranica	stanica	kanta
	spravica	zdravica	bunda
	pelenica	vodenica	krpa
	baštica	koštica	buba
	savanica	tavanica	tabla
	sobica	ubica	kuća
	masnica	mašnica	priča
	kadica	ladica	guma
	najavica	kijavica	kasta
	kravica	krivica	bara
	bananica	pijanica	gužva
	banjica	brnjica	kifla
	pesmica	pesnica	korpa
	rolnica	bolnica	basna
	zabavica	bradavica	palma
	tašnica	košnica	pošta
	sarmica	samica	rana
	saunica	sapunica	tegla
INFLECTIONS			
	bajam	sajam	kidaju
	derem	bagrem	beru
	diktiram	diagram	gadjaju
	jedem	bedem	kluju
	kradem	badem	kažu
	majam	zajam	diraju
	orem	harem	dignu
	pajam	jaram	kupaju
	perem	melem	pasu
	pijem	prijem	dodju
	pletem	sistem	podju
	postim	kostim	dave
	progonim	sinonim	dele
	tonem	fonem	pišu
	udaram	epigram	biraju
	vajam	najam	padaju
	zajam	pojam	bubaju
	zidam	islam	pevaju

Haskins Laboratories Status Report on Speech Research
1993, SR-115/116, 99-112

Visual and Phonological Determinants of Misreadings in a Transparent Orthography*

G. Cossu,†, ‡ D. P. Shankweiler,††† I. Y. Liberman,††† and M. Gugliotta†

Growth of word reading skills was examined in first and second year Italian school children by analysis of the pattern of reading errors. The study was designed to investigate the role of visual vs. phonological similarities as causes of misreadings in a transparent orthography. The selection of reading material was tailored to permit a meaningful cross-language comparison with pre-existing findings on English-speaking children. The results showed that, in Italian as in English, spatially-related errors (such as confusing b and d) constituted a minor proportion of the total errors. Errors on vowel and consonant letters that are not spatially confusable accounted for the greater proportion of the total. Moreover, the co-occurrence of spatial and phonological confusability resulted in appreciably more errors than when either occurred without the other. Vowel position in the syllable had no systematic effect on errors. In beginning readers of Italian, consonant errors outnumbered vowel errors by a wide margin; the reverse pattern was found in previous studies on English-speaking children at the same level of schooling. It is proposed that differences between Italian and English in the phonological structure of the lexicon and in the consistency of grapheme-phoneme correspondences account in large part for the differences in quantity and distribution of the errors.

Unlike the acquisition of spoken language, which develops in the normal child merely through immersion in a speaking environment, learning to read can be a frustrating enterprise for many children. Extracting the linguistic message from seemingly bizarre scribbles may appear to the beginning reader to be an unnatural act (Gough & Hillinger, 1980; Vellutino, 1987). Learning to read in an alphabetic system demands abilities that do not develop automatically from experience with the spoken language. These include: 1) apprehending the letters and their serial arrangement; 2) abstracting the (morphophonemic) units of the linguistic code to which the letter combinations correspond 3) accessing the appropriate lexical representations; 4) integrating the results of orthographic decoding with syntactically-driven parsing operations (Byrne, 1992; Gleitman & Rozin, 1977; Liberman, Shankweiler & Liberman, 1989; Stanovich, 1985). Mastery of these skills requires prolonged instruction and much practice. In view of the special cognitive requirements of reading, it is understandable that many children fail to master and coordinate all the necessary skills. We might expect that a deficiency of a particular skill would be reflected in the kinds of misreadings that occur. Thus, misreadings can be diagnostic of the sources of difficulties.

In most early research on the dyslexic child, developmental dyslexia was viewed as a disorder of the visual aspects of reading (Hinshelwood, 1917). Accordingly, the visuo-spatial properties of letters and words were stressed in contrast to their linguistic functions. A special source of difficulty, on this view, was attributed to the potential right-to-left reversibility of some letter sequences, and to the confusability of letters of

---

The Authors thank Cinzia Avesani, Carol Fowler, Leonard Katz, Alvin M. Liberman, Eric Lundquist, John C. Marshall, Rebecca Treiman, and Mario Vayra for valuable comments on earlier drafts of this paper.

Until the time of her death in July of 1990, Isabelle Y. Liberman contributed importantly to each phase of this research up to the final preparation of the manuscript.

similar shape which differ in orientation, such as "b-d," "p-q," "u-n" (Orton, 1925; 1937).

An unfortunate shortcoming of discussions of reversal errors in the literature is that reversals have been reported selectively in isolation from other aspects of the error pattern. When these errors are viewed in the context of the totality of misreadings, their relative importance tends to diminish. In a study by Liberman and her colleagues, letter confusion and reversal of sequence were found to account for only a small proportion of the errors in oral reading even among very poor readers (Liberman, Shankweiler, Orlando, Harris, & Bell-Berti, 1971; see also: Fisher, Liberman, & Shankweiler, 1978; Shankweiler & Liberman, 1972; Werker, Bryson, & Wassenberg, 1989). Moreover, developmental studies have shown that reversal errors are not peculiar to children with reading difficulties, but can be detected during the normal development of reading skill (Gibson, Gibson, Pick, & Osser, 1962). Finally, there is no clear indication that children who at an early stage tend to confuse letters of similar shape are more likely to remain poor readers than those who do not (Simner, 1982; Mann, Tobin, & Wilson, 1987).

Nonetheless, errors prompted by visual confusability do undoubtedly occur in some children who lag behind in reading. Thus, a possible role for visual confusions as a factor contributing to reading difficulties cannot be dismissed. No research study to date has fully disentangled the relative contribution, and possible interactions, of visual perceptual factors, on the one hand, and linguistic and orthographic factors on the other. The approach we adopt, which, oddly, has apparently not been exploited, is to systematically compare the error pattern in reading the same items presented in upper and lower case script. Since some of the letters, notably the reversible ones, have different shapes in the two scripts, we can directly measure the effects of visual similarity (i.e., reversibility) on frequency of misreading. That was one of the purposes of the present study.

Cross-language variations

Given that earlier research strongly implicates the importance of the nonspatial, linguistic-phonological aspects of the reading process in determining the misreadings that occur, the present study was also designed to identify those aspects of the error pattern that may vary across languages. It has been found, in fact, that the error patterns of beginning readers of English differentially reflect the phonological class of the misread segment and its position within the syllable (Shankweiler & Liberman, 1972). Our point of departure was a discrepancy in the error rate on consonants and vowels. Children learning to read English have consistently shown a higher proportion of errors on vowels than on consonants (Bryson & Werker, 1989; Fisher et al., 1978; Fowler, Liberman & Shankweiler, 1977; Liberman et al., 1971). This effect of category could reflect phonological differences between the two classes of segments and/or differences in the relative difficulty of the spellings of consonants and vowels (Shankweiler & Liberman, 1972).[1]

Analysis of errors with phonologically-controlled materials has also uncovered effects of the position of the target segment within the word: Consonants in the final position of monosyllabic words (and nonwords) are generally more frequently misread than consonants in initial position (Fowler et al., 1977; Liberman et al., 1971). In contrast, the placement of a vowel within the syllable—whether it was the initial, medial or final segment—had no effect on the frequency with which it was misread. It was noted by Fowler et al. (1977) that consonant errors bore a close phonetic relationship to their target phonemes (reflected in distinctive feature similarity), while vowel errors were apparently unrelated to their targets by phonetic feature.

It is significant that the greater propensity to err on vowels may be language-specific, however. For beginning readers of Serbo-Croatian, it has been found that relatively few decoding errors occur, especially among vowels (Ognjenovic, Lukatela, Feldman, & Turvey, 1983). This finding would fit with the transparency of the orthography of Serbo-Croatian, but, alternatively, it may reflect the paucity of vowels in Serbo-Croatian. In regard to the error rates for consonants in syllable-initial and syllable-final position, however, the findings in Serbo-Croatian were consistent with English: Reading final consonants turned out to be less accurate than initial ones.

The work of Ognjenovic et al. illustrates that cross-language comparison can reveal differences and similarities in the error pattern that may have significance for identifying the sources of the problems of learning to read (see Liberman, Liberman, Mattingly & Shankweiler, 1980). Languages that exploit the alphabetic principle vary as to the particular sublexical features that are most directly reflected in word spellings. In some it is almost invariably the phonemic structure that is captured; other orthographies do as

English does, often giving representation to the morphophonemes (Chomsky & Halle, 1968; Venezky, 1970). Owing to the special characteristics of English orthography, especially the propensity to represent morphological structure, the error pattern of children learning to read English may differ significantly from that of beginning readers of languages in which the mode of orthographic representation is more narrowly phonological. Yet up to the present time, a disproportionate share of the research on children's reading problems, including error analyses, has been confined to readers of English. We should not necessarily expect the results of these studies to provide a trustworthy map of the course of learning to read in an orthography that maps rather differently. In this study we focus on one such language, Italian.

Structural differences between Italian and English: Consequences for the beginning reader

Spoken Italian has fewer vowels than English, seven in stressed position and only five in unstressed (Ferrero, Magno-Caldognetto, Vagges, & Lavagnoli, 1978). Moreover, in regard to their acoustic spectra, Italian vowels are highly distinct and nonoverlapping in formant frequencies. Spoken English, on the other hand, has a dozen or more vowels, since the seven basic vowel nuclei are significantly modified by the presence of an off glide (Agard & DiPietro, 1965). Central vowels in English (General American Dialect) show spectral overlap, especially in their reduced form (Peterson & Barney, 1952).

Italian has a relatively shallow phonology, with relatively little morphophonological alternation in comparison to English. In addition, though Italian has a mixed stock of syllable types, it has fewer than half as many different types as English (Carlson, Elenius, Granstrom and Hunnicut, 1985). Moreover, unlike English, which has a predominantly close-syllable structure (e.g., CVC, CVCC, CCVC, etc.), Italian's most frequent syllable form is the open syllable (in sequences such as CVCVC, CVCV, CCVCV, etc.) with relatively few variations (Carlson et al., 1985).

The Italian orthography displays a high degree of transparency, since the alphabetic rendition of the language is based on an almost biunivocal correspondence between grapheme and phoneme. Thus, each of the five vowel letters has only one phonologic rendition in Italian regardless of the context in which it occurs. Similarly, each phoneme generally has an invariant orthographic representation.[2]

English, on the contrary, is represented by a deep orthography. English spellings tend to preserve the complex system of morphophonemic alternations in the language (Chomsky & Halle, 1968). However, no matter how well the orthography of English may comport with the morphological intuitions of the literate user, it would exact a cost from the beginner. Thus, some English spellings (e.g., HEAL-HEALTH) indicate shared morphemes at the expense of consistency in rendering phonological structure.

Having outlined relevant structural differences between the English and Italian languages and their corresponding orthographies, we should list the expected differences in reading and reading-related activities: 1) Because of Italian's open-syllable structure and relatively shallow phonology, learners of Italian should perform better on metalinguistic tasks that tap awareness of phonological segmentation; 2) The error pattern in reading should differ with respect to the relative difficulty of vowels and consonants. Thus, we would expect that, in contrast to English, vowels in Italian would be less often misread than consonants, given their limited number and straightforward orthographic rendition; 3) On the other hand, we would expect that the visual-spatial difficulties based on letter confusability, to the extent that they occur, would cut across differences in language and orthography.

Data pertinent to point 1 are already available. Thus, in a cross-language study of phonological segmentation in Italian-speaking and English-speaking children (Cossu, Shankweiler, Liberman, Katz, & Tola, 1988), we showed that Italian preschool children were able to profit from the simpler syllable structure of their language in performing a metalinguistic task. They proved to be more proficient than their American counterparts in analyzing both the syllabic and phonemic structure of spoken words. Up until now, however, there has been no systematic comparison of the error pattern in beginning readers of the two languages. The present research undertakes to determine whether the differences in language and orthography yield qualitative as well as quantitative differences in misreadings.

Using the methods and results of Liberman et al. (1971) on English-speaking children as a point of departure, we attempted to replicate as closely as possible the tasks and experimental procedures with Italian children. We examined Italian children with respect to: 1) the relative frequencies of visuo-spatial vs. phonological confusions; 2) the distribution of reading errors within the word; 3) the comparative error rate on vowel and consonant segments.

Two experiments on beginning readers of Italian were carried out: in the first, the test words were chosen to represent the principal spelling patterns of consonants and vowels in Italian, and to present maximal opportunities for spatial confusions among the reversible consonants, b,d,g,p,q. In the second experiment, non-words were created to permit the investigation of errors in relation to position of the vowel within the syllable. Each test list was presented in both upper and lower case.

EXPERIMENT 1

Experiment 1 was designed to address two issues: a) the role of visuo-spatial vs. phonological factors in reading errors; b) a cross-language comparison between a transparent and a deep orthography. In order to replicate as closely as possible the conditions of previous studies with English-speaking children (Liberman et al., 1971) a list of 60 words was selected from first and second year reading vocabularies (see Appendix I). Because of peculiarities of each language it was not possible to reproduce all the characteristics of the test words of the earlier study. For example, the material prepared by Liberman et al. (1971) distinguished between sight words, non- sight words and word-forming reversals. In Italian, by contrast, neither the sight-word - non-sight word distinction can be made, nor is it possible to create more than a few reversible letter strings that form a different word when read from right to left.

Method

The Italian subjects were 70 school children (35 males and 35 females) randomly selected from first and second year classes of an elementary school in the northern Italian city of Parma. We restricted our selection to the earliest school years, because pilot work had shown that by the end of the second school year most Italian children make very few errors in decoding. In the school selected for our study, reading is taught eclectically. None of the children experienced anything approximating a pure phonic or a pure whole language approach. All children with known or suspected history of brain damage were excluded from the experimental sample, as well as those children with clinically-evident language impairment, visual or auditory deficits, or behavioral disorders. As shown in Table 1, all the children were within the normal range of intelligence, according to the Verbal Scale of the WISC.

We adopted the following criteria for the Italian materials: 1) all the words were bisyllabic; 2) the list included each of the 15 Italian consonants (in a CV sequence); 3) each consonant appeared twice in the first syllable and twice in the second one, for a total of 60 words. The list included 44 CVCV words; in 16 cases, however, (due to the limitations of children's vocabularies) we were forced to include CCV sequences in the non-critical syllable (appended to the CV sequence of the target).

In order to examine the role of spatial features of the letter set, the same word list was presented twice, once in upper case and once in lower case. The interval between the two testing sessions was one week and the order of presentation for the upper vs. the lower case list was counterbalanced. Each word was printed on a separate index card in upper and lower case. The cards for each list were placed face down in front of the subject and were turned over one by one by the examiner. The children were asked to read each word as it was presented and to give their best guess if they were unsure. They were tested individually during the middle of the school year. Their responses were recorded on magnetic tape and phonetically transcribed by the examiner.

TABLE 1. *Mean age and IQ of Italian school children.*

Group	Age (months) Mean	IQ (WISC Verbal) Mean
Year 1 (n=35)	82.03 (75-87)	102.06 (81-129)
Year 2 (n=35)	94.28 (90-98)	109.09 (80-134)

The findings from the Italian-speaking subjects were compared with data from English-speaking beginning readers from the United States who had been studied, using the same methods of investigation, by two of the authors (see Liberman et al., 1971 for details). Although the two samples are as similar as was practically attainable demographically, in type of school and in the instructional approach taken to reading, it is patent that such matching can only be approximate. Within these constraints, there were differences in the criteria by which subjects were selected. The American children (N = 18) constituted the lower third of a second year elementary school population defined by score on the experimental word reading test. The Italian sample, as noted, was a random selection from first and second year students in the targeted schools. These differences in selection procedure do not permit unequivocal quantitative comparisons across national groups. However, in view of the fact that many aspects of the error pattern in English-speaking learners have been shown to be quite stable across wide differences in level of attainment, it is reasonable to assume that valid comparisons can be made regarding the relative frequencies of different categories of misreadings (see Fowler et al., 1977; Shankweiler & Liberman, 1972).

Scoring of reading errors was based on the following criteria:

1: Reversal of Sequence (RS) was scored when a word, or part of a word was read from right to left (e.g.: palo, as "aplo," "olap," or "lopa").

2: Reversal of Orientation (RO) was scored when b,d,p,q and g were mutually confused. In order to check the effect of visual vs. phonological factors, the same criteria were adopted for the scoring of these letters in upper case.

3: Consonant Errors (CE), unless otherwise specified, comprise all the errors on consonants, other than the RO category. Consonant substitution, deletion, or insertion were therefore included under this head.

4: Vowel Errors (VE), included all vowel substitutions, deletions or insertions.

5: Complex Errors (CXE) were scored when more than two errors occurred in one word. (e.g., when cinque /ʧinkwe/ [five] was read as /kina/.

Results

The number of phonemic segments misread in the 60-item word list was tallied for each subject. In the first year, errors per subject averaged 9.1 for upper case and 12.5 for lower case. In the second year the corresponding means are 3.4 and 4.4, respectively. As expected, second year students made fewer errors on the reading test than first year students. In each year, words printed in lower case characters turned out to be slightly more difficult to read than in those in upper case. The low rate of errors, even in first year children, testifies to the rapid acquisition of decoding skills in Italian beginning readers. With regard to the qualitative aspects of the error pattern, Figure 1 shows the distribution of errors among the error categories for years one and two combined.

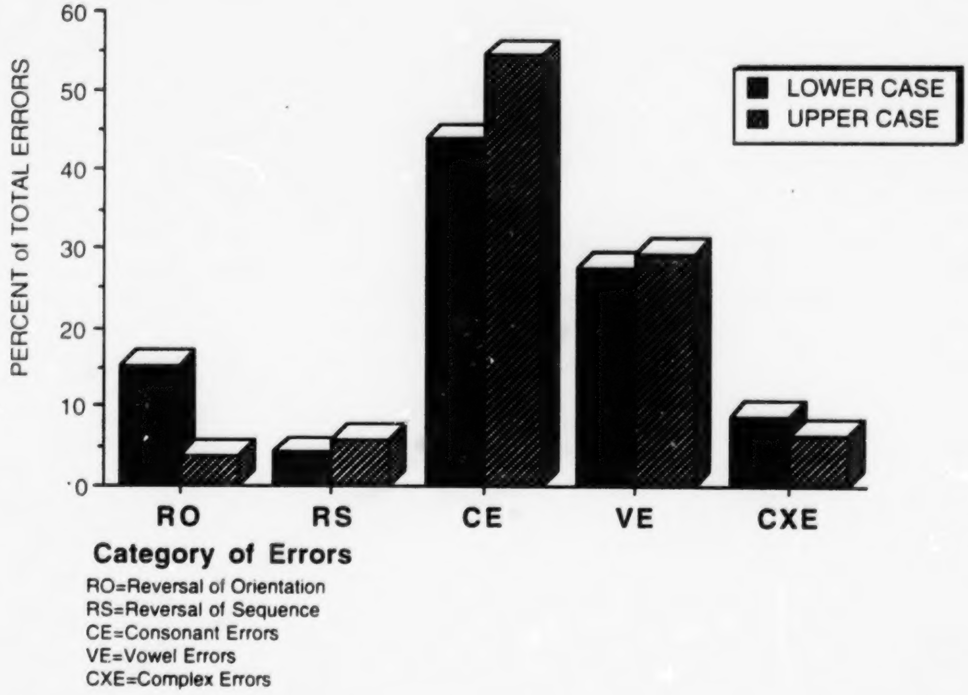


Figure 1. Mean percent of reading errors in upper and lower case tabulated by category for 1st and 2nd year Italian school children.

In this figure, RO error rates are given also for the upper case counterparts of the reversible letters though these, of course, are not reversible. On inspection of Figure 1 we note that consonant errors other than reversals (CE) together with vowel errors (VE) accounted for the bulk of the errors, whereas the proportions of RS, RO and complex errors (CXE) are comparatively low. Two findings stand out: first, consonant errors predominate over vowel errors (a point to which we return in the discussion). Second, within each error category (other than RO errors) the difference between the error rates engendered by upper and lower case is small.

An analysis of variance was carried out in which errors from all categories (total error) was the dependent variable and year in school and letter case were the independent variables. School year [F(1,32) = 9.27, p <.003] and letter case [F(1,32) = 16.51, p <.0001] are each significant factors, together with their interaction [F(1,32) =5.20, p <.03]. Though it was small absolutely, the significant effect of case, and the significant interaction confirms that lower case presentation was indeed somewhat more difficult for the first year students.

The effect of spatial reversibility

Mean frequencies of errors by category, and corresponding percentages based on opportunities, are presented in Table 2. As for letter reversal errors, the table shows that these are almost entirely confined to the first-year children. Right-to-left readings of whole words (RS errors) occurred infrequently even among the younger children.

In order to probe the influence of letter case on confusability, we rank-ordered the 60 words of the test list according to frequency of RO error, separately for upper and lower case, for first and second year students. The rank order of difficulty for words presented in each case agrees closely, although lower case presentation elicited more errors in 54 of the 60 words. A rank correlation analysis shows that upper- and lower-case performance is significantly correlated in both the first and second year (rho = .73, z = 7.14, p > .0001 and rho = .53, z = 4.50, p > .0001, respectively).

Since there was a small excess of errors in the lower case format, we conducted a further analysis by category to locate the source of the discrepancy. For this analysis, we subdivided the nonreversible consonant errors (the CE category) into errors of substitution (CE1) and errors of addition and deletion (CE2). A Wilcoxon test was carried out for each subcategory of errors, separately for first and second year groups, to evaluate the significance of each difference. At each school year, there were significantly more lower-case than upper-case errors in the set b,d,g,p,q (first year, p <. 001; second year, p <. 03). As expected, the largest discrepancy between upper- and lower-case emerged for the RO subcategory. In the combined sample, RO errors accounted for 15.3% of the total, whereas the equivalent set in upper case (which are not spatially confusable) yields only 3.8% of errors.

TABLE 2. *Frequencies of error by category for upper and lower case and percentages (in parentheses) of opportunities [in brackets].*

	FIRST YEAR		SECOND YEAR	
	Upper	Lower	Upper	Lower
RO Reversal of Orientation [1260]	16 (1.26)	74 (5.87)	1 (0.08)	17 (1.35)
RS Reversal of Sequence [2100]	11 (0.52)	13 (0.61)	15 (0.71)	13 (0.62)
CE 1 Consonant Substitution [4200]	93 (2.21)	101 (2.40)	31 (0.74)	47 (1.12)
CE 2 Consonant Addition and Deletion [4200]	84 (2.0)	85 (2.02)	31 (0.74)	27 (0.64)
VE Vowel Errors [4200]	90 (2.14)	124 (2.95)	39 (0.93)	40 (0.95)
CXE Complex Errors [2100]	25 (1.19)	43 (1.02)	3 (0.14)	10 (0.48)

It is worth noting that within the set of spatially confusable consonant characters, there are correlated phonological similarities. For example, all are stops that share distinctive features in common. In view of this, it is not surprising that b was often misread for d (36 times), since the phonemes b/ and /d/ share all their features except place of production. Similar examples can be cited for upper case characters, where G /ʤ/ was misread for C /ʧ/ 16 times. However, u was never misread for n and the reverse happened only once, in spite of their figural resemblance. Yet, two letters showing little visual similarity, but high phonological similarity (d and t) were confused 11 times in lower case and 11 times in upper case. Thus, though spatial similarity may contribute to reading difficulties for beginners, it is not sufficient by itself to elicit reading errors. With the exception of vowel errors in year 1 (where lower case errors were also more frequent [$p <. 03$]), no other difference attributable to case reached significance. Thus, the excess of errors on lower case is attributable chiefly to the letters b,d,p,g and q. The pattern of errors on this set is also notably different in lower case and upper case. In lower case, the error entailed a confusion within the set on 85% of presentations; on only 15% was a reversible target letter misread as a nonreversible letter. By contrast, in upper-case presentation of the corresponding letters there were few confusions within the set. In order to display the pattern of confusions among the reversible lower-case consonants, the data were arrayed in a confusion matrix. Only the data for the first year children are presented because so few errors occurred during the second year. As shown in Table 3, there is scant tendency for reciprocity in confusions between members of reversible letter pairs. For example, in the first year b is misread for d 36 times while d is misread for b only eight times. Asymmetry is also found for q, which is misread 13 times as p, whereas p is never misread as q.

TABLE 3. *Matrix of confusion among reversible lower case consonants for first and second school year combined.*

		SUBSTITUTION				
		b	d	g	p	q
TARGET	b	--	36	3	3	0
	d	8	--	2	1	0
	g	0	0	--	0	0
	p	4	1	0	--	0
	q	0	0	4	13	--

Table 4 shows correlations among the error categories, aggregated over subjects, for year 1 and year 2 Italian children combined. It is notable that word reversal errors (RS) are uncorrelated with reversal errors that involve individual letters (RO). The latter, on the other hand, tend to co-occur with errors that are clearly nonspatial. Thus, there seems to be no justification for the common practice of grouping the RS and RO together. Lack of association between RO and RS was also noted in the study of reading errors in American school children by Liberman et al. (1971). The cardinal importance of language-related factors in misreadings is clearly evident when we examine the role that phonological category plays in the error pattern.

TABLE 4. *Correlations among categories of reading errors in combined Italian groups*.

	R.S.	R.O.	C.E.	V.E.
R.S.	--	-.04 $p < .36$	.20 $p <. 04$*	.27 $p < .01$*
R.O.		--	.60 $p < .001$*	.49 $p < .001$*
C.E.			--	.65 $p < .001$*
V.E.				--

Cross-language comparison

Strictly speaking, we cannot make a direct comparison on quantity of errors between the present findings on Italian children and those of Liberman et al. (1971) on English-speaking children. Differences in the criteria for subject selection, and the impossibility of equating the instructional content in the two countries mean that the consistently lower error rates in the Italian children cannot be given an unequivocal interpretation. Notwithstanding this, great care was exercised in creating the test materials in this study so as to provide a comparable frame for comparison of the distribution of errors in relation to letter-case and phonological category.

We reasoned that the poorer second-grade readers (from Liberman et al., 1971) and the average first-grade readers (from the Italian sample) were comparable groups for assessing similarities and differences in decoding errors. In order to support our assumption that the distribution of errors within the word does not depend crucially on the level of ability, we compared the Italian good and poor readers from the first and second school year. For this comparison we selected the reading list from experiment 2 (the "vowel reading test"), since it was made up of monosyllables, and thus resembled, in so far as possible with Italian materials the list presented to the American children.

Moreover, we reasoned that nonwords, more likely than real words, would evoke qualitative differences in the reading strategies of reader groups differing in ability level, if such differences existed.

Error rates were ranked for Italian children in the first and second school year in both upper- and lower-case reading conditions. The nine best and poorest readers were selected from each school class. The mean error rate in the first year was 4.4 (S.D.=3.2) for poor readers and 0.4 (S.D.=1.01) for good readers. In the second school year, the corresponding means were 2.1 (S.D.=1.36) for poor readers and 0.11 (S.D. =0.33) for good readers. The results were submitted to a four-way ANOVA with two between (grade and reading ability) and two within factors (letter case and position of vowel). As expected, grade [F(1,32)=11.92, *p* <002] and reading level [F(1,32)= 39.57, *p* <.001] were significant, as well as their interaction [F(1,32)= 9.74, *p* <.004]. The effect of letter case was nonsignificant, as was the grade by letter-case interaction, the reading-level by letter-case interaction, and the grade by reading-level by letter-case interaction. In addition, the vowel position factor was nonsignificant as well as the related interactions: grade by position, reading level by position and grade by reading level by position, and all higher-order interactions.

The analysis revealed that overall error patterns did not differ across ability levels. Neither letter case nor vowel position was treated in a qualitatively different way by good and poor readers. Thus, the results of this analysis suggest that the reading strategies adopted by average beginning readers and older poor readers were essentially the same, in agreement with findings on English-speaking readers who differed in ability level (cf., Fowler et al., 1977, and Bryson & Werker, 1989). Therefore, the comparison between Italian beginning readers and American (comparatively older) second grade poor readers is valid for the purpose of comparing qualitative features of the error pattern across languages.

In Figure 2, we show the distribution of errors among the four categories as proportions of total error for the first year Italian sample and the American second year poor readers. First, letter reversals (RO), account for a similar proportion of the total error in each group (17% for Italians and 15% for Americans). The confusions among the reversible letters are presented in Table 5 in the form of a confusion matrix which includes the data for both the Italian and American samples.

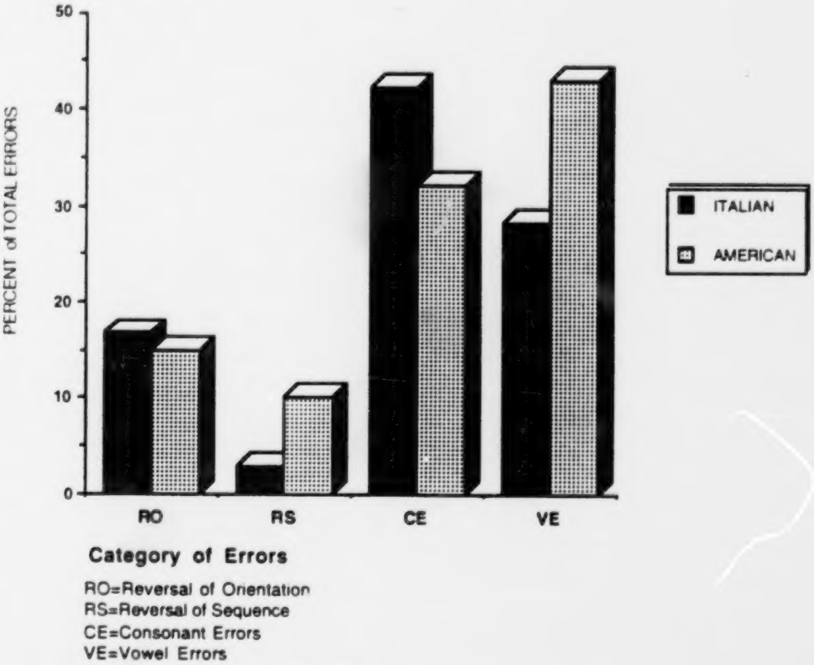


Figure 2. Comparison of reading errors according to category in 1st year Italian and 2nd year English-speaking children: Percentages of total error.

Table 5. *Confusions among reversible letters in American and Italian children (Percentages based on Opportunities).*

	OBTAINED				
PRESENTED	b	d	p	g	
b	--	10.2	13.7	0.3	A
		14.2	1.2	1.2	I
d	10.1	--	1.7	0.3	A
	3.2		0	0.8	I
p	9.1	0.4	--	0.7	A
	1.2	0.3		0	I
g	1.3	1.3	1.3	--	A
	0	0	0		I

A= American (data taken from Liberman et al., 1971)
I = Italian

Both groups confined their errors on the set b,d,p,q,g, chiefly to confusion among b,d, and p. The one notable difference is that the American children, but not the Italians, tended to reverse in each direction, i.e., vertically (e.g., b,p and p,b) as well as horizontally. In neither language group does the distribution of responses conform to the relative frequencies of occurrence of these letters in samples of text (in the case of English, we used published tables by Mayzner & Tresselt, 1965; for Italian, we made our own letter count). Right-to-left sequence reversals (RS) occur very infrequently in Italian, probably because they rarely form meaningful words. Nonreversible consonant and vowel errors together account for most of the total error in both language groups, but the relative frequencies are reversed in the two samples, being nearly mirror images of one another. The Italians, unlike the English-speaking learners, but like the Serbo-Croatian beginning readers, misread consonants more often than vowels.

EXPERIMENT 2

The purpose of Experiment 2 was to test for the effect on reading errors of varying the position of the vowel within the syllable. Because the Italian language has predominantly an open-syllable structure, its lexicon contains relatively few monosyllabic words with syllable-final consonants. Therefore, nonwords are required to make a satisfactory test for the effect of position of a target vowel letter.

Method

A list of monosyllabic nonwords was made, each comprising a vowel and two consonants. Each of the five vowels appeared twice in each position (initial, medial and final), for a total of 30 nonwords (see Appendix II). These materials were presented to the same groups of Italian first- and second-year students who served as subjects in Experiment 1. As in Experiment 1, upper-case and lower-case forms were prepared. These were presented one week apart in counterbalanced order. The child's task was to read each nonword aloud. Errors were tallied on both consonants and vowels. Furthermore, errors were classified according to the position of the vowel (initial, medial, final) within the word, and the type case (upper/lower). However, no attempt was made to assess the effect of position of the consonant because relatively few words in the lexicon of Italian contain closed syllables.

Results

As expected, children from the second school year outperformed the younger group in reading the monosyllabic nonwords [$F(1.32) = 11.92$, $p < .002$]. The effect of case in this experiment did not approach significance in either first or second year students. Therefore, we pooled the data for upper and lower case: the first year students misread a mean of 9.4 segments per subject and the second year students misread 3.6 per subject. Comparing these results with those of Exp 1, we find that the error rates for nonwords and real words are similar.

Of central interest is the discrepancy between vowel and consonant errors. With nonwords, as with words (Exp. 1), vowel reading was more accurate than consonant reading. By examining the percent of errors as a function of opportunities (Table 6), we see that at both age levels and in both upper and lower case, vowels represent a lower proportion of the total error than consonants. Indeed, vowel errors occurred at a rate of well under 1% among the second year students.

TABLE 6. *Frequencies of errors percentaged as a function of opportunities on vowels in nonwords.*

SCHOOL YEAR	CASE	CONSONANTS n=[2100]	VOWELS n=[1050]
1st	Upper	135 (6.4%)	37 (3.5%)
	Lower	130 (6.2%)	27 (2.6%)
2nd	Upper	46 (2.1%)	3 (0.2%)
	Lower	69 (3.2%)	8 (0.7%)

Possible reasons for the discrepancies between the language groups in the relative frequency of vowel and consonant errors are considered in the following discussion.

The test of the effect of position of the vowel within the syllable yielded a null result for the combined school years. That is, when the vowel was placed in initial, medial and final position within the syllable, the errors did not vary in any systematic way. In this respect, the data agree with findings on English-speaking children for both words (Fowler et al., 1977; Liberman et al., 1971) and nonwords (Bryson & Werker, 1989).

GENERAL DISCUSSION

In discussing the findings of the two experiments, we consider first the data on beginning readers of Italian and then interpret the findings in relation to existing data on beginning readers of English and other languages.

Firstly, with regard to quantity of misreadings, Italian beginning readers made strikingly few decoding errors, especially after the first year at school. Secondly, within each category of errors except RO, the difference between the error rate in upper and lower case is minimal. It is apparent, however, that within the subset of spatially reversible characters (b,d,p,q,g), visual similarity does contribute to reading difficulties for beginners, indicated by the significant excess of errors on these consonants in lower case. The frequencies of these errors diminished greatly in the second year, but the effect of case remains significant. Thirdly, the greater proportion of the total error is attributable to nonreversible consonant letters and vowel letters. Thus, with the exception of the reversible consonant set noted above, the error pattern reflects not the spatial characteristics of the misread letters, but their functions within the linguistic system and its orthographic representation.

In the case of the spatially-reversible consonants, it is notable that reversibility is not ordinarily sufficient by itself to elicit reading errors. Scrutiny of the confusions shows that only when visual similarity is associated with phonological similarity of the corresponding phoneme (as in the case of b and d, where these letters ordinarily represent phonemes that differ by only one distinctive feature) were misreadings apt to occur. In a like fashion, a low degree of phonological similarity can forestall misreadings even when high visual confusability exists (e.g.,, u and n, and F and E are rarely confused). Moreover, even when a letter pair presents good figural contrast, high phonological similarity (e.g., d and t) may lead to a high rate of confusions. Further examination of the confusions reveals that the matrix of substitution errors was similar in both upper and lower case modality. In upper case, substitutions were more likely to occur when minimal phonetic distance is combined with visual confusability. Thus, in the first and second year combined, G is misread for C 17 times, whereas L, E and F were never confused. But D was misread for T 11 times, while B was never misread for R, or vice versa. Similarly, in lower-case presentation, z was misread for s seven times, due to the fact that these letters represent a phonetically-similar segment by means of a visually similar shape. Visual similarity alone is clearly not sufficient to elicit substitution errors, as evident from the absence of substitution errors between the vowel u and the consonant n. Hence, the two sources of similarity, though they interact, produce unequal effects: Phonological similarity tends to override visual similarity.

Taken together, these findings confirm the indications of previous studies of beginning readers of English in pointing to the central relevance of language-related structural (phonological and morphological) factors in reading acquisition, with visuo-spatial factors being relegated to a relatively minor role[3] (Fowler et al., 1977; Liberman et al., 1971; Shankweiler & Liberman, 1972).

The distribution of errors in beginning readers of Italian reflects the phonological structure of Italian and the transparent orthography that renders it. Thus, as expected, some aspects of the error pattern contrast with what has been found in readers of English. For example, the small number of vowels in Italian, their nonoverlapping acoustic spectra and the fact that each tends to be consistently represented by the same letter—all cooperate to minimize the occurrence of vowel errors. Similarly, the greater number of consonants in Italian (relative to English), and the complexities of their orthographic rendition in some cases (as for /ʧ/ vs. /k/, for example) together may account for the preponderance of consonant errors in Italian. For both upper and lower case, consonant errors exceeded vowel errors by a wide margin.

The stability of the excess of consonant errors (relative to vowel errors) in Italian children's misreadings is confirmed by the results of Experiment 2, which provides an independent demonstration of this pattern with non-word test materials. Here, too, the findings on the Italian children contrast strikingly with those of their English-speaking peers: Vowel errors as a function of opportunity constitute a low proportion of the total.

In comparing misreadings of Italian-speaking and English-speaking beginning readers, we find both similarities and differences. In each orthography, visual-spatial confusability of characters is of roughly the same magnitude as measured by their proportions of the total error. But when we examine errors as percentages of opportunities for error of each kind, a different picture emerges: letter reversals loom larger in relation to non-spatial errors in the Italian children, whereas in English-speaking beginning readers, the reverse is true. This is not because reversal errors are absolutely more frequent in Italian than in English: Indeed, they are less frequent, but their importance seems greater because other errors are comparatively fewer. As for sequence errors, in both the Italian and American beginning readers, right-to-left sequence reversals (RS) occur relatively infrequently. Sequence reversals amount to 4.3% and 10% of the total in the Italian and the American children, respectively. As we noted, their lower rate of occurrence in Italian probably reflects the paucity of reversible words in that language.

The study was not intended to yield a direct comparison between American and Italian groups with regard to quantity of errors, but it is impossible not to be impressed by the disparities in this regard, though interpretation must be tempered by the differences between the two studies in subject selection. Notwithstanding that, the results showed that the first-year Italian children, though a year younger and exposed to half as much schooling, were much more accurate, making far fewer errors (as a function of opportunity) than the American children. These findings are compatible with the inference that the orthographies of Italian and English pose somewhat different problems for a beginning reader. The Italian children appear to master the relevant decoding skills much more rapidly than their American age mates, as also suggested by findings of Cossu et al. (1988) and Lindgren, DeRenzi, and Richman (1985).[4]

We have pointed to some cross-language differences that are surely relevant: the transparency of the Italian orthography, the relatively shallow phonology with fewer syllable types and morphophonological alternations, as well as the limited number of vowels, all of which are acoustically well-spaced. These characteristics would likely cooperate to minimize the obstacles for the Italian beginning reader. It is apparent that the structural differences between the two languages and their orthographies find expression in the contrasting patterns of consonant and vowel errors. A three-way comparison of error data among beginning readers of English, Serbo-Croatian and Italian is useful for elucidating the patterns. English and Serbo-Croatian, though sharing a closed syllable structure, differ in the size and complexity of the vowel set and in the structure of their orthographies. The Serbo-Croatian vowel set contains only five vowels, and, as in Italian, their acoustic spectra (formant frequencies) are distinct and nonoverlapping (Ognjenovic et al., 1983). In keeping with these phonological and phonetic differences, the orthographies of English and Serbo-Croatian also differ in the nature of the correspondences between letters and word structure: The Serbo-Croatian orthography is highly phonographic, while English is more strongly morphologically influenced. It is significant that in the study by Ognjenovic et al., beginning readers of Serbo-Croatian, made fewer errors on vowels than on consonants, in contrast to their American counterparts, but like the Italians in the present study.

Like English-speaking learners studied by Fowler et al. (1977), the Italians showed no significant position effect for vowels in Experiment 2. The further question of presence or absence of phonological similarity of the vowel substitutions could not be addressed in this study, however, because even the first-year Italian children made so few vowel errors that no analysis in terms of distinctive features could be carried out. As noted earlier, the open-syllable structure that prevails in Italian precludes a comparison of error rates on initial and final consonants. However, in both Serbo-Croatian and English, each of which has closed-syllable structure, the same position effect on consonants was observed: beginning readers misread more consonants in syllable-final position than in syllable- initial position (Ognjenovic et al., 1983).

Script systems in use in different languages display diverse means for adapting the alphabetic principle to the structural peculiarities of the language so as to minimize arbitrariness, redundancy and ambiguity (Klima, 1972). The facts regarding their diversity led to the expectation that orthographies would draw somewhat unequally upon a range of cognitive abilities, and that the cognitive demands associated with different systems should be reflected in the pattern of reading errors (see Liberman et al., 1980). In the main, the findings

confirm these expectations, and also the expectation that cross-language comparisons of the progress of beginning readers have potential value for delineating more precisely the set of specific skills required for mastery of each orthography. The present study of reading errors in beginning readers of Italian, together with the comparative findings on English and Serbo-Croatian, points to significant differences in reading processes that are associated with these three orthographies that share the alphabetic principle. At the same time, the findings point to the existence of common problems in learning to read in an alphabetic system.

REFERENCES

Agard, F. B., & Di Pietro, R. J. (1965). *The sounds of English and Italian*. Chicago: The University of Chicago Press

Bryson, S. E., & Werker, J. F. (1989). Toward understanding the problem in severely disabled children. Part. I: Vowel Errors. *Applied Psycholinguistics, 10*, 1-12.

Byrne, B. (1992). Studies in the Acquisition Procedure for Reading: Rationale, hypotheses, and data. In P. B. Gough, L. C. Ehri, & R. Treiman (Eds.), *Reading acquisition* (pp. 1-34). Hillsdale, NJ: Lawrence Erlbaum Associates.

Carlsson, R., Elenius, K., Granstrom, C., & Hunnicut, S. (1985). Phonetic and orthographic properties of the basic vocabulary of five European languages. *Quarterly Report (KTH Speech Transmission Laboratory, Stockholm), 1*, 63-94.

Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. New York: Harper and Row.

Cossu, G., Shankweiler, D., Liberman, I. Y. Katz, L., & Tola, G. (1988). Awareness of phonological segments and reading ability in Italian children. *Applied Psycholinguistics, 9*, 1-16.

Ferrero, F. E., Magno-Caldognetto, Vagges, K., & Lavagnoli, C. (1978) Some acoustic characteristics of the Italian vowels. *Journal of Italian Linguistics, 3*, 87-96.

Fischer, F. W., Liberman, I. Y. & Shankweiler, D. P. (1978). Reading reversal and developmental dyslexia: a further study. *Cortex, 14*, 496-510.

Fowler, C., Liberman, I. Y., & Shankweiler, D. (1977). On interpreting the error pattern in beginning reading. *Language and Speech, 20*, 162-173.

Gibson, E. J., Gibson, J. J., Pick, D. & Osser, R. A. (1962). Developmental study of the discrimination of letter-like forms. *Journal of Comparative and Physiological Psychology, 55*, 897-506.

Gleitman, L. R. and Rozin, P. (1977). The structure and acquisition of reading: Relation between orthography and the structure of language. In A. S. Reber & D. L. Scarborough (Eds.), *Toward a psychology of reading* (pp. 1-53). Hillsdale: Erlbaum.

Gough, P. B., & Hillinger, M. L. 1980. Learning to read: An unnatural act. *Bulletin of the Orton Society, 30*, 179-196.

Hinshelwood, J. (1917). *Congenital word blindness*. London: H. K. Lewis & Co.

Klima, E. S. (1972). How alphabets might reflect language. In J. F. Kavanagh & I. Mattingly (Eds.), *Language by ear and by eye* (pp. 57-80). Cambridge: MIT Press.

Liberman, I. Y., Liberman, A. M., Mattingly, I., & Shankweiler, D. (1980). Orthography and the beginning reader. In J. Kavanagh & Venezsky, R. (Eds.), *Orthography, reading and dyslexia* (pp. 137-154). Baltimore: University Park Press.

Liberman, I. Y., Shankweiler, D. S., & Liberman, A. M. (1989). The alphabetic principle and learning to read. In D. Shankweiler & I. Y. Liberman (Eds.), *Phonology and reading disability: Solving the reading puzzle* (pp. 1-33). Ann Arbor, MI: University of Michigan Press.

Liberman, I. Y., Shankweiler, D., Orlando, C., Harris, K. S., & Bell-Berti, F.(1971). Letter confusions and reversals of sequence in the beginning reader: Implication for Orton's theory of Developmental Dyslexia. *Cortex, 7*, 127-142.

Lindgren, S. D., DeRenzi, E., & Richman, L. C. (1985). Cross-national comparisons of developmental dyslexia in Italy and the United States. *Child Development, 65*, 1404-1417.

Mann, V. A., Tobin, P., & Wilson, R. (1987). Measuring phonological awareness through the invented spellings of kindergarten children. *Merrill Palmer Quarterly, 33*, 365-392.

Mayzner, M. S., & Tresselt, M. E. (1965). Tables of single-letter and diagram frequency counts for various word-length and letter-position combinations. *Psychonomic Monograph Supplements, Vol. 1*, No. 2, 13-18.

Ognjenovic, V., Lukatela, G., Feldman, L., & Turvey, M. T. (1983). Misreadings by beginning readers of Serbo-Croatian. *Quarterly Journal of Experimental Psychology, 35*, 581-615.

Orton, S. T. (1925). "Word blindness" in school children. Archives of Neurology and Psychiatry, 14, 581-615.

Orton, S. T. (1937). Reading, writing and speech problems in children. New York: Norton.

Peterson, G. E., & Barney, H. L. (1952) Control methods used in the study of the vowels. The Journal of the Acoustic Society of America, 24, 175-184.

Shankweiler, D. P., & Liberman, I. Y. (1972). Misreading: a search for cause. In Kavanagh and Mattingly (Eds.) Reading by ear and by eye (pp. 293-317). Cambridge: MIT Press.

Simner, M. (1982). Printing errors in kindergarten and the prediction of academic performance. *Journal of Learning Disabilities, 15*, 155-159.

Stanovich, K. (1985). Explaining the variance in reading ability in terms of psychological processes: what have we learned? *Annals of Dyslexia, 35*, 67-96.

Vellutino, F. (1987). Dyslexia. Scientific American, 256, 34-41.

Venezky, R. (1970). The structure of English orthography. The Hague: Mouton.

Werker, J. F., Bryson, S. E., & Wassenberg, K. (1989). Toward understanding the problem in severely disabled readers Part II: Consonant errors. Applied Psycholinguistics, 10, 13-30.

FOOTNOTES

[*] A slightly different version will appear in *Reading & Writing*.

[†] Servizio di Neuropsichiatria Infantile, University of Parma.

[‡] Neuropsychology Unit, Dept. of Clinical Neurology, Radcliffe Infirmary, University of Oxford.

[†††] Haskins Laboratories and University of Connecticut, Storrs.

[1] In English, the vowels tend to have multiple spellings, whereas consonants display greater consistency in grapheme-phoneme correspondences (Venezky, 1970).

[2] An exception is the voiceless velars /k/ and /g/ which have a different spelling depending on the following vowel. The letters [c] and [g], when followed by the vowels a, o and u are rendered as voiceless stop consonants /k/ and /g/ respectively. When followed by the vowels e and i, they are rendered as affricates /ʧ/ and /ʤ/.

[3] It is possible, of course, (as suggested in Fischer et al., 1978) that children who evince widespread impairment of higher visual processes may face a special obstacle in coping with the visual processing requirements of reading.

[4] See Gough and Hillinger (1980) for further information on the rate of learning decoding skills in reading by English-speaking children.

APPENDIX I

WORD READING LIST FOR CONSONANTS

(upper and lower case)

Target	1st syll.		2nd syll.		Target	1st syll.		2nd syll.	
[B]	baci	*kisses*	ruba	*s/he steals*	[P]	palo	*pole*	rapa	*turnip*
	bocca	*mouth*	cibo	*food*		pollo	*chicken*	rospo	*toad*
[C]	cena	*dinner*	brace	*ember*	[Q]	quadro	*picture*	acqua	*water*
	carte	*cards*	buca	*hole*		questo	*this*	cinque	*five*
[D]	dado	*die*	corda	*rope*	[R]	riso	*rise*	ieri	*yesterday*
	doni	*presents*	grido	*scream*		rami	*branches*	sera	*evening*
[F]	faro	*lighthouse*	buffa	*funny*	[S]	seme	*seed*	scuse	*apologies*
	filo	*threat*	gufi	*owls*		sole	*sun*	casa	*house*
[G]	gara	*competition*	strega	*witch*	[T]	tino	*vat*	denti	*teeth*
	gelo	*frost*	urge	*it urges*		tana	*lair*	trota	*trout*
[L]	lato	*side*	gola	*throat*	[V]	vigna	*vineyard*	rovi	*bramble*
	lepre	*hare*	sale	*salt*		vaso	*pot*	riva	*shore*
[M]	mela	*apple*	rame	*copper*	[Z]	zanna	*fang*	pizza	*pizza*
	mano	*hand*	crema	*cream*		zoppo	*flame*	quarzo	*quartz*
[N]	nero	*black*	pane	*bread*					
	naso	*nose*	vena	*vein*					

APPENDIX II

WORD READING LIST FOR VOWELS

(UPPER AND LOWER CASE)

[A]	art	ant	sab	car	spa	cra
[E]	erp	est	set	der	pre	sme
[I]	int	inc	mit	cip	cri	spi
[O]	ont	ort	cor	fon	sto	pro
[U]	urt	umb	sup	cub	stu	tru

Haskins Laboratories Status Report on Speech Research
1993, SR-115/116, 113-125

Phonological Computation and Missing Vowels: Mapping Lexical Involvement in Reading*

Ram Frost[†]

The role of assembled versus addressed phonology in reading was investigated by examining the size of the minimal phonological unit that is recovered in the reading process. Readers named words in unpointed Hebrew that had many or few missing vowels in their printed forms. Naming latencies were monotonically related to the number of missing vowels. Missing vowels had no effects on lexical decision latencies. These results support a strong phonological model of naming and suggest that even in deep orthographies phonology is not retrieved from the mental lexicon as a holistic lexical unit but is initially computed by applying letter-to-phoneme computation rules. The partial phonological representation is shaped and completed through top-down activation.

Although the process of reading acquisition ultimately involves the extraction of meaning from print, there is a fairly general agreement that at some stage this process requires the recovery of phonologic information from the orthographic structure. How exactly the printed form is converted into phonology is a topic for current debate. Two possible mechanisms have been suggested to account for the reading process. The first mechanism assembles phonology from print by applying a set of conversion rules (or through weighted connections in a neural network) that transform letters, letter clusters, or graphemes into phonemes or phonemic clusters. The assembly of phonology in this case is a computational process that involves a set of transformations that connect minimal orthographic and minimal phonologic units (letters and phonemes in the case of alphabetic orthographies like English; letters and syllables in the case of syllabic orthographies like Japanese; graphemes and morphemes in the case of logographic orthographies like Chinese).

The second mechanism involves a direct mapping of whole-word orthographic units into whole-word phonologic units. The complete phonologic structure of the printed word is then addressed by its orthographic form and retrieved as a whole from the mental lexicon. Thus, in contrast to assembled phonology, addressed phonology does *not* involve any computation at the subword unit level, but is derived from straightforward connections between the printed and the spoken representations of a word.

The relative use of addressed versus assembled phonology in naming has been the focus of heated debates because it bears on an old but fundamental issue in the reading literature: the speed and efficiency of visual-orthographic encoding in visual word recognition (see Katz & Frost, 1992, for a review). What is usually labeled *the visual encoding hypothesis* assumes that regardless of the type of orthography, it is usually more efficient to visually access the lexicon and retrieve from it the complete phonologic structure of the printed word rather than to assemble it using prelexical conversion rules. This is because the visual encoding hypothesis posits that at least for high-frequency words visual encoding is fastest and involves minimal cognitive resources. Moreover, the visual encoding hypothesis assumes that, once the lexicon is accessed, the process of retrieving the phonologic information from it does

---

This work was supported in part by National Institute of Child Health and Human Development Grant HD-01994 to Haskins Laboratories. I am indebted to Rona Segev for her help in conducting the experiments, and to Ken Forster, Bruno Repp, Alexander Pollatsek, and two anonymous reviewers for their comments on earlier drafts of this paper.

not involve significant cognitive effort (e.g., Baluch & Besner, 1991; Besner & Smith, 1992; Seidenberg, 1985; Tabossi & Laghi, 1992). In contrast to the visual encoding hypothesis the *phonological hypothesis* suggests that the default operation of the cognitive system in word recognition is the use of prelexical rather than addressed phonology. The basic argument of the phonological hypothesis is that all writing systems are phonological in nature and their primary aim is to convey phonologic structures, i.e. words, regardless of the graphemic structure adopted by each system (see De Francis, 1989; Mattingly, 1992, for a discussion). The computation of phonological structures from print is, therefore, a primary function of the system. The phonological hypothesis suggests that if readers can successfully assemble a prelexical phonological representation from print, then at least in the naming task, it will be used first. The easier it is to generate a prelexical representation the more often it will be used. For example, if an orthography is shallow in that it has direct and consistent correspondences between letters and phonology, then readers of this orthography will be able to utilize these correspondences for naming and will use minimal resources in the process (see Katz & Frost, 1992, for a discussion).

The phonological hypothesis is supported by studies showing extensive phonologic recoding in shallow orthographies (e.g., Feldman & Turvey, 1983; Frost, Katz, & Bentin, 1987; Katz & Feldman, 1981, 1983; Turvey, Feldman, & Lukatela, 1984; and see Carello, Turvey, & Lukatela, 1992, for a review), and by findings showing that readers in shallow orthographies strategically prefer prelexical phonological assembly over the retrieval of phonological information from the lexicon following visual access (Frost, 1994). Moreover, the phonological hypothesis gains support also from increasing evidence that prelexical phonology is used by readers of deeper orthographies, like English (e.g., Perfetti, Bell, & Delaney, 1988; Perfetti, Zhang, & Berent 1992; Van Orden, 1987; Van Orden, Johnston, & Hale, 1988).

Two versions of the phonological hypothesis can be distinguished. The weak version views the generation of phonological information from print as a process that may involve, in principle, *both* addressed and assembled phonology. According to this hypothesis the relative clarity of mapping between orthography and phonology determines how exactly phonology is derived from print. To cast it in activation terms, the weak hypothesis proposes that there are computations at the level of subword units (e.g., letter-to-phoneme), but there are also direct connections of whole-word orthographic units and whole-word phonologic units, allowing whole-word orthographic units to directly activate whole-word phonologic units regardless of the computation at the prelexical level. What determines the final outcome of such a "race" is the ease with which prelexical processing may be achieved. Thus, although the default of the system is to assemble phonology from print, the prelexical computation process could be bypassed, and phonology may be entirely addressed rather than assembled when the orthography represents phonology in a complex way, and the relations between graphemes and phonemes are inconsistent and opaque (Frost & Katz, 1989; Frost, 1994). Note that such complexity may be found both within writing systems (i.e., irregularly spelled words), and as a factor distinguishing between writing systems, so-called deep and shallow orthographies (e.g., Frost et al., 1987).

On the other hand, the *strong* phonological hypothesis argues that a model for generating phonology from print does not need to assume connections between whole-word orthographic units and whole-word phonologic units, and that phonology is *always* assembled. Thus, in any alphabetic orthography, the initial process of recovering phonologic information from print necessarily involves a computation of graphemes into phonemes. The computed phonological representation may well be affected by lexical knowledge if it is poor or incomplete. However, the strong phonological hypothesis denies that lexical effects in pronunciation result from visual access of the lexicon and retrieval of phonologic information of the whole word from it (i.e., from direct activation of whole-word phonologic units by whole-word orthographic units). Rather, the mandatory process of transforming letter clusters into phonemic clusters is said to be interactively affected by lexical knowledge through top-down activation. In a nutshell, the strong phonological hypothesis does not make use of the notion of "addressed phonology" at all, since phonology is never entirely addressed, but always computed (e.g., Carello et al., 1992; Lukatela & Turvey, 1990; Lukatela, Turvey, Feldman, Carello, & Katz, 1989; Seidenberg & McClelland, 1989; Van Orden, Pennington, & Stone, 1990).

An example of a strong phonological model is the one offered by Turvey and his colleagues to represent reading in bi-alphabetical writing systems like Serbo-Croatian (e.g., Lukatela et al., 1989). The architecture of this model allows the reader a fast computation of phonology in both Roman and Cyrillic writing systems even though they share letters representing different phonemes (e.g., "B" representing the phoneme /b/ in Roman script but /v/ in Cyrillic script). The model specifies how the Roman and the Cyrillic graphemic units in the Serbo-Croatian reader's lexicon are connected to phonemic units, without allowing any direct links between whole printed words units and whole spoken words units. Orthographic ambiguity is resolved entirely by interactive processes between the word unit level and the phoneme unit level within the phonologic lexicon. Carello et al. (1992) argue that this framework can account for naming in all alphabetic orthographies (see also Carello, Turvey, & Lukatela, in press).

Providing empirical evidence to distinguish between the weak and the strong versions of the phonological hypothesis is not a simple task because both models predict prelexical as well as lexical effects in naming. Note that the difference between "addressed" and "lexically shaped" phonology is a unit-size difference. That is, the distinction between the two versions lies only in the way in which word pronunciation is obtained-as a whole-word unit following visual lookup, or as a top-down shaping of prelexical phonological computation of subword units. The aim of the present study was to address this theoretical distinction with a methodology that can be easily implemented in Hebrew.

In Hebrew, letters represent mostly consonants while most of the vowels can optionally be superimposed on the consonants as diacritical marks ("points"). The diacritical marks, however, are omitted from most reading material, and can be found only in poetry, children's literature, and religious scriptures. Since different vowels may be inserted into the same string of consonants to form different words or nonwords, Hebrew unpointed print cannot specify a unique phonological unit. Therefore, a printed consonant string is always phonologically ambiguous and often represents more than one word, each with a different meaning. Some vowels, however, (mainly /o/, /u/, /i/) may be represented in print not only by points but also by letters (See Baluch & Besner, 1991, for a similar characteristic of Persian). These letters are not always used, and are often considered optional by the writer.[1] When they are used, however, the complete phonologic structure of unpointed Hebrew words may be uniquely specified by the print. For example, the word "מיתון" (/mitun/-meaning *recession*) contains two vowels, each of them represented by a letter (ו,י). Note that the phonologic structure of such a word can be assembled almost as easily as it can be assembled in pointed print.[2] Because some words in unpointed Hebrew include vowel letters and some do not, printed words differ in their level of phonological ambiguity. The following theoretical construct aims to characterize the nature of this ambiguity.

DEGREES OF FREEDOM

When readers of Hebrew are presented with an unpointed printed word that can be meaningfully pronounced in only one way (i.e., lexically unambiguous word), they face the problem of assigning to the letter string the correct vowel configuration, so as to interpret or pronounce the printed word correctly. This process of filling in the missing vowels characterizes the reading of almost any word in unpointed Hebrew, even if it lexically unequivocal. The concept of *Degrees of Freedom* (DF) represents the amount of ambiguity involved in this process. Consider the following computational rule:

> *Every letter that represents a consonant which may potentially take a vowel, adds one degree of freedom to the reading process, whereas any consonant letter that is disambiguated by a following vowel letter, does not.*

If, for example, a consonant letter is followed by another consonant letter, the initial letter can be pronounced, in principle, with any vowel (or with a silent vowel) and contributes, according to the above definition one DF to the reading task. If, on the other hand, a consonant letter is followed by a vowel letter, the cluster represented by the two letters is a phonologically unequivocal syllable, which does not add any DFs to the reading process. Final letters are in most cases not followed by any vowel, and do not add DFs to the reading process. In a nutshell, the number of DFs a word contains, refers to the number of vowels not represented by letters, and consequently, reflects the amount of missing phonological information that is necessary for correct pronunciation of this word.

It is important to note that although the above rule allows an easy computation of the number of DFs a printed word contains, this number only approximates the level of phonological ambiguity faced by the reader. First, our computation procedure does not take into account the number of possible vowels that each consonant can take, or their relative probabilities. Second, it does not consider constraints on permissible vowel patterns in Hebrew or on possible combinations of vowels and consonants allowed by the Hebrew morphology (for a discussion of Hebrew morphology, see Bentin & Frost, in press; Frost & Bentin, 1992; Shimron, 1993). The number of DFs presents, therefore, only coarse-grained calculation of the amount of ambiguity each word poses to the reader. Nevertheless, for the purpose of the present study, this measure seemed sufficient.

Degrees of freedom and naming time

In the present context, DFs are of theoretical importance only when phonology is computed and assembled at the level of sub-word units. If the phonological representation of a printed word is addressed following visual access and retrieved as a unit from the mental lexicon, at least for lexically unambiguous words, the overall number of DFs computed from the individual letters should not play a significant role in naming. What characterizes the process of addressing phonology from the lexicon is the direct connection between a holistic orthographic cluster and a holistic phonological structure. If the visual word pattern directly and unequivocally addresses one spoken word, the phonemic information conveyed by single letters should not affect naming. Thus, for unambiguous words, the number of missing vowels in unpointed Hebrew print should matter only if reading involves some computation at the letter to phoneme level.

The strong phonological hypothesis makes specific predictions concerning the effect of DFs on naming latencies. Not only does it predict that DFs should affect naming performance, but it predicts that this effect should be monotonic. Thus, for a given number of letters in a printed word, the more DFs these letters represent, the slower naming should be. This is because the bottom-up process of computing phonology from print recovers only partial phonological information; the consonantal phonemic cluster. The vowel information necessary for correct pronunciation must be filled in through top-down activation from the word unit level to the phoneme unit level. However, the more phonemes have to be filled through this interactive process, the more impoverished is the computed prelexical representation, the slower the buildup of activation within the system, and consequently the slower the naming latencies will be.

Monitoring the effect of DFs on naming performance would, therefore, constitute a critical test concerning the validity of the weak and the strong phonological hypothesis. Showing that DFs are good predictors of naming latencies would suggest that prelexical phonological computation occurs in the pronunciation task. On the other hand, showing that DFs do not affect naming performance would support the claim that phonology is mainly addressed in Hebrew, rather than assembled, as the weak phonological hypothesis would predict. Previous studies have shown that the reader of unpointed Hebrew relies extensively on orthographic recoding in word recognition (Bentin & Frost, 1987; Frost, 1992; Frost & Bentin, 1992a, 1992b; Frost et al., 1987). For example, Bentin and Frost (1987) have shown that lexical decisions for unpointed Hebrew ambiguous words were faster than lexical decisions to *either* of the disambiguated pointed alternatives. This outcome suggested that lexical decisions in unpointed Hebrew were based on the early recognition of the orthographic structure that was shared by the phonological and semantic alternatives. Thus, a demonstration of prelexical computation in an orthography as deep as unpointed Hebrew, would provide significant evidence in support of the strong phonological hypothesis.

EXPERIMENT 1

Experiment 1 measured naming latencies for a corpus of 256 unpointed words differing in their DF values and their frequency. Both DF and frequency were collapsed into high and low levels. The aim of the experiment was to examine whether words with a large number of DFs will be named slower than words with a small number of DFs, and whether this effect interacts with frequency.

Method

Subjects. The subjects were 42 undergraduate students at the Hebrew University, all native speakers of Hebrew, who participated in the experiment for course credit or for payment.

Stimuli and design. The stimuli consisted of 256 Hebrew words that were three to five letters long, and contained two or three syllables with five to eight phonemes. All words were unambiguous and could be pronounced as only one meaningful word. Words were classified as high- or low-frequency words; and as having a large or a small number of DFs. This created four groups of words, 64 words in each group. DFs were calculated following the rules described above. For different word lengths, the corpus included a high-DF word and a low-DF word that could be either high-frequency or low-frequency. For example, a four-letter word could have three DFs if the four letters consisted of four consonants (the final letter almost never takes a vowel), or one DF if the four letters included a vowel that disambiguated a CVC cluster. Both the high-DF and low-DF words could be either frequent or nonfrequent, etc. Examples of various Hebrew words with a large or small number of DFs are presented in Figure 1.

In the absence of a reliable frequency count in Hebrew, the subjective frequency of each word was estimated by 50 undergraduate students who rated the frequency of each word on a 7-point scale, from very infrequent (1) to very frequent (7). The rated frequencies were averaged across all 50 judges. The average frequencies for high-DF words were 4.82 for frequent words and 2.46 for nonfrequent words. The average frequencies for low-DF words were 4.97 for frequent words and 2.58 for nonfrequent words.[3]

Procedure and apparatus: The stimuli were presented on a Macintosh II computer screen in a bold Hebrew font, size 24 (5mm). Subjects were tested individually in a dimly lighted room. They sat 70 cm from the screen so that the stimuli subtended a horizontal visual angle of 4 degrees on the average. Naming latencies were monitored by a Mura-DX 118 microphone connected to a voice key. Each experiment started with 16 practice trials, which were followed by the 256 experimental trials presented in two blocks. The intertrial interval was 2.5 sec.

Results

Naming latencies were averaged across subjects for high- and low-frequency words with high- and low-DFs. Within each subject/condition combination, RTs that were outside a range of 2 SDs from the respective mean were excluded, and the mean was recalculated. Outliers accounted for less than 5% of all responses. This procedure was repeated in all the experiments of the present study. Because nonwords were not included in the experiment, the overall percentage of errors (mainly wrong pronunciations) was quite small (1%) and did not allow a reliable analysis. The results are presented in Table 1.

DFs affected naming latencies; high-DF words were slower to name than low-DF words. The statistical significance of the results was assessed by an analysis of variance (ANOVA) across subjects (F_1) and across stimuli (F_2), with the main factors of DFs and frequency. The main effect of DFs was significant ($F_1(1,41) = 27.6$, $MS_e = 230$, $p < 0.001$, $F_2(1,252) = 7.2$, $MS_e = 1363$, $p < 0.007$), as was the main effect of frequency ($F_1(1,41) = 139$, $MS_e = 191$, $p < 0.001$, $F_2(1,252) = 28.8$, $MS_e = 1363$, $p < 0.001$). The two-way interaction was significant in the subjects analysis ($F_1(1,41) = 9.1$, $MS_e = 164$, $p < 0.004$), but not in the stimuli analysis ($F = 1.9$).

One possible source of the obtained DF effect is the number of words having zero DFs. Because the phonological structure of these words could have been computed entirely prelexically it is possible that all of the DF effect has emerged because these words were contrasted with words having one or more DFs. Note that the strong phonological hypothesis predicts a *monotonic* effect of DFs, and not merely a difference between phonologically opaque and phonologically transparent words. In order to verify that the DF effect did not result just from fast RTs to zero DFs words and similar RTs to all the other words having one DF or more, only words with four letters were examined. As can be seen in Figure 1, all four-letter words in the corpus had either one or three DFs, but never zero DFs. Thus, even in the low-DF condition these words were not entirely phonologically transparent. The results of this post-hoc procedure are presented in Table 2. As can be seen, the same, and even greater, advantage of low-DF words over high-DF words was obtained. Thus, it is clear that the overall DF effect did not emerge just from the inclusion of zero DFs words.

	High -DF Words	Low-DF Words
Printed Form	קבלן - KBLN (3DFs)	נזיר - NZIR (1DF)
Pronunciation and meaning:	/kablan/ - ("contractor")	/nazir/ - ("monk")
Printed form:	פסנתר - PSNTR (4 DFs)	תינוק - TINOK (0 DF)
Pronunciation and meaning:	/psanter/ - ("piano")	/tinok/ - ("baby")

Figure 1. **Examples of high- and low-DF Hebrew words.**

Table 1. *Naming latencies (and SDs) for low- and high-frequency words with low- and high-DFs. Words are unpointed.*

	Low-DF Words	High-DF Words
Low-frequency words	533	552
	(37)	(41)
High-frequency words	514	521
	(32)	(37)
Mean RTs	524	537

Table 2. *Naming latencies for low- and high-frequency four-letter words having one or three DFs. Words are unpointed.*

	1-DF Words	3-DF Words
Low-frequency words	532	556
	(37)	(37)
High-frequency words	510	523
	(29)	(41)
Mean RTs	521	540

The effect of number of phonemes on RTs was examined as well, because on the average, low-DFs words have fewer phonemes than high-DF words when the number of letters is kept constant.[4] Thus, it was important to make sure that the number of phonemes *perse* did not affect naming latencies. The mean RTs for words having four or five phonemes was 530 ms, whereas the mean RTs for words having six to eight phonemes was 532 ms, suggesting that the number of phonemes in itself did not affect naming time.

Discussion

The results of Experiment 1 suggest that DFs affect naming time. When the number of letters was kept constant, the more DFs were contained in a printed word, the longer were the naming latencies. This outcome suggests that the phonological structure of the printed words was not retrieved as a unit from the mental lexicon following visual access, but was assembled via letter-to-phoneme correspondences. When vowels are missing in the orthographic representation, only a partial phonological representation can be computed by the assembly process. This partial representation can be completed only through a top-down shaping process that involves lexical knowledge. As the number of missing vowels increased, this interactive process slowed down, resulting in slower naming latencies.

It is possible to gain some insight into the involvement of the mental lexicon in pronunciation by examining the frequency effect. The results suggest a reliable frequency effect across DFs, supporting the notion of lexical involvement in naming. However, if our hypothesized computation procedure is correct, a significant frequency effect should appear only when there is some ambiguity in the printed word, that is, only with DFs greater than zero. In contrast, words having zero DFs should not show any frequency effect, or should show it to a much lesser extent. This is because zero DFs words contain all the phonemic information in print, and their phonological structure can be assembled prelexically without lexical contribution. An analysis of the frequency effect across DFs supports these predictions. Table 3 depicts the frequency effects for each DF level. There was a fairly strong frequency effect for all DFs greater than zero, but it became small and nonsignificant (6 milliseconds only) for words having zero DFs.

Table 3. *Frequency effects in naming with words having zero to four DFs.*

	0-DF	1-DF	2-DF	3-DF	4-DF
High-frequency	529	510	525	523	513
Low-frequency	535	532	549	556	546
Frequency effect	6	22	24	33	33

Another point of interest concerning word frequency is the size of the DF effect for high- and for low-frequency words. The results are not unequivocal. Tables 1 and 2 suggest that DFs had a greater effect for low-frequency words than for high-frequency words. This interaction, however, was significant only in the subject analysis. One possible problem with the item analysis is that frequency was made into a dichotomous variable. Therefore, a more robust test of the interaction was carried out by treating the frequency ratings as a continuous variable which served as an RT predictor separately for high- and for low-DF words. A regression analysis revealed that the slopes of the two regression lines were not significantly different ($r = 0.3$ and $r = 0.41$ for high- and low-DF words respectively, Z = 0.76). Thus, the results seem to suggest that DFs affect naming latencies for frequent and nonfrequent words in a similar manner. Whether the size of the effect changes with word frequency is unclear.

EXPERIMENT 2

The aim of Experiment 2 was to provide a baseline for the process of phonological assembly. If naming latencies are affected by the number of missing vowels in the printed word, then the effect of DFs should disappear in pointed print. Pointed Hebrew print disambiguates the consonantal structure by providing the missing vowels in the form of diacritical marks. When the vowel marks are printed, they do not allow any degrees of freedom in reading each of the consonants, and all word are treated as having zero DFs. In Experiment 2, subjects named the same words as in Experiment 1, but all words were fully pointed. The purpose of the experiment was to demonstrate that in this condition DFs will not affect naming time.

Methods

Subjects. The subjects were 42 undergraduate students at the Hebrew University, all native speakers of Hebrew, who participated in the experiment for course credit or for payment. None of the subjects participated in Experiment 1.

The stimuli, design and procedure were identical to those employed in Experiment 1 with the only difference that all stimuli were pointed.

Results

Naming latencies were averaged across subjects for high- and low-frequency words with high- and low-DFs. As in Experiment 1, outliers accounted for less than 5% of all responses. The results are presented in Table 4. DFs had no effect on naming latencies (503 ms for *both* High- and Low-DFs). There was a frequency effect but it was much smaller (14 ms) than the effect obtained in Experiment 1 with unpointed print (27 ms). The overall percentage of errors was less than 1%.

Table 4. *Naming latencies (and SDs) for low- and high-frequency words with low- and high-DFs. Words are pointed.*

	Low-DF Words	High-DF Words
Low-frequency words	508	512
	(36)	(34)
High-frequency words	498	494
	(36)	(34)
Mean RTs	503	503

The statistical significance of the results was assessed by an analysis of variance (ANOVA) across subjects (F_1) and across stimuli (F_2), with the factors of DFs and frequency. Only the effect of frequency was significant, $F_1(1,41) = 49$, $MS_e = 163$, $p < 0.001$; $F_2(1,252) = 9.8$, $MS_e = 1245$, $p < 0.001$). The interaction was significant in the subject analysis ($F_1(1,41) = 12.5$, $MS_e = 75$, $p < 0.001$, but not in the item analysis ($F_2 = 1.1$).

Discussion

The results of Experiment 2 confirm that it is indeed the number of missing vowels that affected naming latencies in Experiment 1. When words are pointed they become phonologically transparent and each word contains practically zero DFs. The addition of vowel marks eliminated the main effect of DF, and reduced the frequency effect considerably. This suggests that in most cases subjects assembled the pointed words' phonology prelexically, using simple letter-to-phoneme conversion rules. These results conform with previous studies in Hebrew showing prelexical strategies of naming in pointed Hebrew (Frost, 1994). Overall, naming latencies in Experiment 2 were faster than in Experiment 1. This outcome is accordance with various studies showing faster naming performance in pointed than in unpointed Hebrew (e.g., Frost, 1994; see also Shimron, 1993). While there was no main effect of DF in Experiment 2, the interaction of DF and frequency was significant in the subject analysis. This could suggest that DF had a more deleterious effect for low-frequency words than for high-frequency words. However, given the instability of the effect this possibility should be treated with caution.

EXPERIMENT 3

The aim of Experiment 3 was to map the effect of DFs on lexical decision for pointed and unpointed words. Previous studies have shown that lexical decisions in Hebrew are based on the recognition of the orthographic structure and are made prior to a complete phonological analysis of the printed word (Bentin & Frost, 1987; Frost & Bentin, 1992b; Frost & Kampf, 1993; Frost, 1994). If lexical decisions do not involve a deep phonological analysis of the printed word, then DFs should not affect decision latencies for both pointed and unpointed words. Whether a letter cluster contains several missing vowels or none, should not affect response time.

Method

Subjects. Sixty undergraduate students at the Hebrew University, all native speakers of Hebrew, participated in the experiment for course credit or for payment. Thirty of them were assigned to the unpointed condition and the other 30 were assigned to the pointed condition. None of the subjects participated in the previous experiments.

Stimuli and design: The stimuli consisted of the same word corpus employed in the naming experiments with the addition of 256 nonwords. Nonwords were created by altering randomly one or two letters of high- or low-frequency real words that were not employed in the experiment. The nonwords were all pronounceable and did not violate the phonotactic rules of Hebrew. The 512 stimuli were divided into two lists, containing 128 words and 128 nonwords each. Half of the subjects in each condition (pointed or unpointed) received one list and half the other list, randomly. The procedure and apparatus were identical to those employed in Experiment 1 and 2 with the only difference that subjects conveyed their decision by pressing a "yes" or a "no" response key. The dominant hand was always used for the "yes" response.

Results

RTs in the different experimental conditions for unpointed and pointed print are presented in Table 5. Again, outliers accounted for less than 5% of all responses. DFs had no effect on lexical decisions in pointed as well as unpointed print. Overall, lexical decision latencies in pointed and unpointed print were very similar.

Table 5. *Lexical decision latencies (and SDs) for low- and high frequency words with low- and high- DFs.*

	POINTED		UNPOINTED	
	Low-DF Words	High-DF Words	Low-DF Words	High-DF Words
Low frequency words	598 (70)	604 (79)	594 (70)	600 (77)
High-frequency words	537 (58)	535 (60)	536 (65)	533 (67)
Mean RTs	568	570	565	567

Similar to the procedure of Experiments 1 and 2, separate analyses were performed on the pointed and unpointed data. The effect of DFs was not significant in both the pointed and the unpointed conditions (F_1, F_2 <1.0). The effect of frequency was significant in both the pointed condition ($F_1(1,29) = 217$, $MS_e = 582$, $p < 0.001$; $F_2(1,252) = 144$, $MS_e = 973$, $p < 0.001$), and the unpointed condition ($F_1(1,29) = 294$, $MS_e = 405$, $p < 0.001$, $F_2(1,252) = 146$, $MS_e = 816$, $p < 0.001$). The two-way interaction was not significant in both the pointed condition ($F_1 = 1.3$, $F_2 = 1.4$), and the unpointed condition ($F_1 = 3.2$, $MS_e = 162$, $p < 0.08$, $F_2 < 1.0$).

Discussion

The results of the lexical decision task confirm that DFs affect performance only when a phonological representation has to be constructed from the print. Several studies have repeatedly shown that lexical decisions are given prior to a deep phonologic analysis of the printed word (e.g., Bentin & Frost, 1987; Frost & Bentin, 1992a; Frost, 1994; and see Frost & Bentin, 1992a, for a review). These studies would predict, therefore, that DFs will not affect lexical decision time. The similar results for pointed and unpointed stimuli are in accordance with a previous study by Koriat (1984) who showed almost identical lexical decision latencies for pointed and unpointed print. In a subsequent study, however, Koriat (1985) found that the presentation of vowel marks had some beneficial effect on lexical decisions for low-frequency words. This evidence, however, was inconclusive. The present data seem to fit better his initial results (Koriat, 1984). The Results of Experiment 3 suggest that DFs are not confounded with factors affecting lexical access or lexical search. Rather, they are relevant only to the recovery of phonology from print.

GENERAL DISCUSSION

The aim of the present study was to examine the role of assembled versus addressed phonology in naming using a novel methodology. The two routes for generating a phonological representation from print are often differentiated in terms of lexical involvement (or lack of it) in naming. Therefore, previous studies have monitored the extent of lexical contribution to correct pronunciation by measuring semantic priming and frequency effects (e.g., Baluch & Besner, 1991; Frost, 1994; Frost et al., 1987). However, the theoretical distinction between assembled and addressed phonology may be based on a different criterion which relates to the *size* of the minimal phonological unit that is recovered in the reading process. When the phonological structure of the printed word is lexically "addressed", it is retrieved as a whole unit from the lexicon following visual access. In contrast, when it is assembled, it is recovered segment by segment through a process of prelexical conversion of letter or letter clusters into phonemes or phonemic clusters. The present study aimed to examine the size of the recovered phonological units in naming by manipulating ambiguity at the letter level.

Unpointed Hebrew provides a unique opportunity to assess the effect of letter ambiguity on pronunciation. This is because each unpointed consonant in Hebrew represents a phonological puzzle to the reader concerning the exact vowel that should follow this consonant. The assessment of ambiguity in the study was somewhat simplified. Each letter slot was treated categorically, by merely assessing whether it added to the overall ambiguity score or not. Obviously, the permissible word patterns in Hebrew constrain the possible vowels that each consonant can take, thereby affecting the level of complexity of each discrete puzzle. Thus, it is possible that the actual contribution of a missing vowel slot to the overall ambiguity score was higher or lower than the contribution of its neighboring vowel slots. Nevertheless, the DF score assigned to each word reflected, to a close approximation, the amount of missing phonological information that was necessary for successful assembly.

DFs allow, therefore, a critical test of two contrasting hypotheses concerning word naming. Does the printed word undergo a process of phonologic computation at the subword unit level, or is the word's phonology retrieved as a holistic unit following a lexical lookup? The words employed in the present study were phonologically and semantically unambiguous at the word level. That is, their orthographic structure pointed to only one lexical entry. This entry contained but one phonological representation and one semantic meaning. Thus, if only the word level is examined, these words did not present to the reader any form of lexical ambiguity. Their phonology could be, in principle, unequivocally retrieved following lexical lookup. The concept of DFs relates only to the ambiguity at the level of letters-to-phonemes conversion. It is exactly this feature which provides the ability to test the weak and the strong phonological hypotheses. If the initial phase of generating a phonological representation from print entails computation at the subword unit levels, then DFs should affect this process. The more ambiguity has to be resolved at the subword level, the longer should be the process of generating a complete phonological representation. If, on the other hand, the orthographic structure is used to access the lexicon visually and retrieve the printed word's phonology following a lexical lookup, ambiguity at the letter level should not affect this process.

The results of Experiment 1 provide significant support for the strong phonological hypothesis. DFs affected naming latencies when both frequency and word length were kept constant. Words with a larger number of DFs took longer to pronounce than words with a smaller number of DFs. This effect was not restricted to a comparison between completely transparent words (having zero DFs) and opaque words, but persisted within opaque words which differed in the number of DFs they contained. The monotonical effect of missing vowels cannot be easily accommodated by a model that considers naming as the result of mapping entire orthographic structures into holistic phonologic structures. The data of Experiment 1 thus suggest that the phonologic representation of the printed words was computed piecemeal rather than retrieved holistically.

Experiments 2 and 3 reinforce this conclusion by providing two independent baselines. Because different words were employed in the high-DF and the low-DF conditions, it was necessary to ensure that the DF effect did not emerge from trivial differences between word samples. In Experiment 2 the words of Experiment 1 were presented in their pointed form. Thus, the only difference between Experiment 1 and 2 was that all words became zero DFs words. The results of Experiment 2 show that when the words were pointed, the main effect of DFs disappeared and RTs to the two word samples were virtually identical. This outcome confirmed that it was indeed the differential ambiguity of the unpointed stimuli that has caused the DF effect in Experiment 1, and not other possible factors related to the stimuli employed in the different experimental conditions.

Similar conclusions arise from the results with lexical decision in Experiment 3. DFs, in general, allow a powerful test of the hypothesis that phonological recoding occurs in lexical decision. Previous studies in Hebrew have established that lexical decisions in Hebrew do not involve a deep phonological analysis of the printed word, but are based on the shallow recognition of letter strings that may represent several different words with different meanings (Bentin & Frost, 1987; Frost, 1992; Frost, 1994; Frost & Bentin, 1992a, 1992b, Koriat, 1984; and see Shimron, 1993 for a review). The results of Experiment 3 confirmed, therefore, that detailed phonologic recoding is not necessary for lexical decisions in Hebrew, and DFs play a role only when a full phonological representation is needed, as is the case in the naming task. The similar, almost identical RTs for low- and high-DF words again suggest that these words have a similar lexical status. Thus, Experiment 3 further supports the contention that the different naming latencies obtained in Experiment 1 were only due to the amount of ambiguity at the letter level as measured by the DF analysis.

An important outcome of Experiment 1 concerns the contribution of lexical factors to the assembly of phonology as revealed by the frequency effect. There was a strong effect of frequency on naming latencies for DFs greater than zero. This result suggests that lexical involvement occurred whenever a complete phonological representation could not be assembled only using letter-to-phoneme conversion rules. When the printed words were completely transparent, word frequency did not have a strong effect on naming latencies, suggesting that phonology was assembled with minimal lexical contribution. However, when some phonological ambiguity was present in the letter string, lexical involvement was immediately apparent.

However, a more important conclusion concerning the frequency effect is its co-occurrence with the effect of DFs. Previous studies that examined the relative use of addressed versus assembled phonology have distinguished between the two routes by monitoring the existence of lexical involvement in naming. Lexical factors like semantic priming or word frequency were taken as evidence for getting the word's phonology through a process of lexical lookup and not through prelexical computation (e.g., Baluch & Besner, 1992, Frost, 1994). Thus, according to the classical dual-route view, phonology can be either lexical or assembled and the two routes for obtaining it are independent (e.g., Paap, Noel, & Johansen, 1992). The present study suggests that the process of generating a phonological representation may involve simultaneously both prelexical and lexical processing. Thus, if the orthography is not extremely shallow, both processes come into play. By this view, the two routes are not functionally independent but interact to allow a correct pronunciation.

A model that can accommodate these results is an interactive model that views the process of generating phonology from print as a process of converting letters or letter clusters into phonemes or syllables. This process, however, in most cases cannot compute a complete and accurate phonological representation. Thus, the output of the computation process is shaped by top-down lexical knowledge that inserts missing phonemic information like the missing vowels of unpointed Hebrew, or fills in the correct pronunciation of irregular words in English. Such a process is exemplified in Figure 2. The figure depicts the phases of naming a high-DF Hebrew words like "לפתן" (LFTN, pronounced /liftan/, meaning *desert*). LFTN has three missing vowel slots. The initial phase of getting its phonology is a computation process that transforms the consonantal information into phonemes and creates an incomplete phonological representation. The vowels are inserted during or following this process, whether serially or in parallel, by top-down lexical shaping. This provides a complete phonological representation that allows the correct pronunciation. Hence, the complete phonological representation of LFTN is not retrieved from the lexicon as a holistic unit following visual access caused by the four letters. Rather, it involves both an assembly process and a lexical contribution. This model is very similar in nature to the model proposed by Lukatela et al. (1989) to account for reading in Serbo-Croatian, and is in accordance with the strong phonological hypothesis. The importance of the results in unpointed Hebrew is that the demonstration of an assembly process in such a deep orthography provides significant support for the strong phonological view.

Although the Model in Figure 2 describes the computation of phonology in Hebrew, its general framework can serve to account for pronunciation in English or any alphabetic orthography. Note that the ambiguity faced by the Hebrew reader is different from that faced by English speakers. In Hebrew, the mapping of letters into phonemes is fairly consistent, and phonological ambiguity results from missing phonemic information in print. In English, on the other hand, the letters represent all of the word's phonemes but in an inconsistent manner. However, recent results in English show a striking similarity to the results obtained in the present study. Using a backward masking paradigm, Berent and Perfetti (1993) have shown that the phonological representation of English CVC words is not lexically addressed but computed in two processing cycles with different time courses.

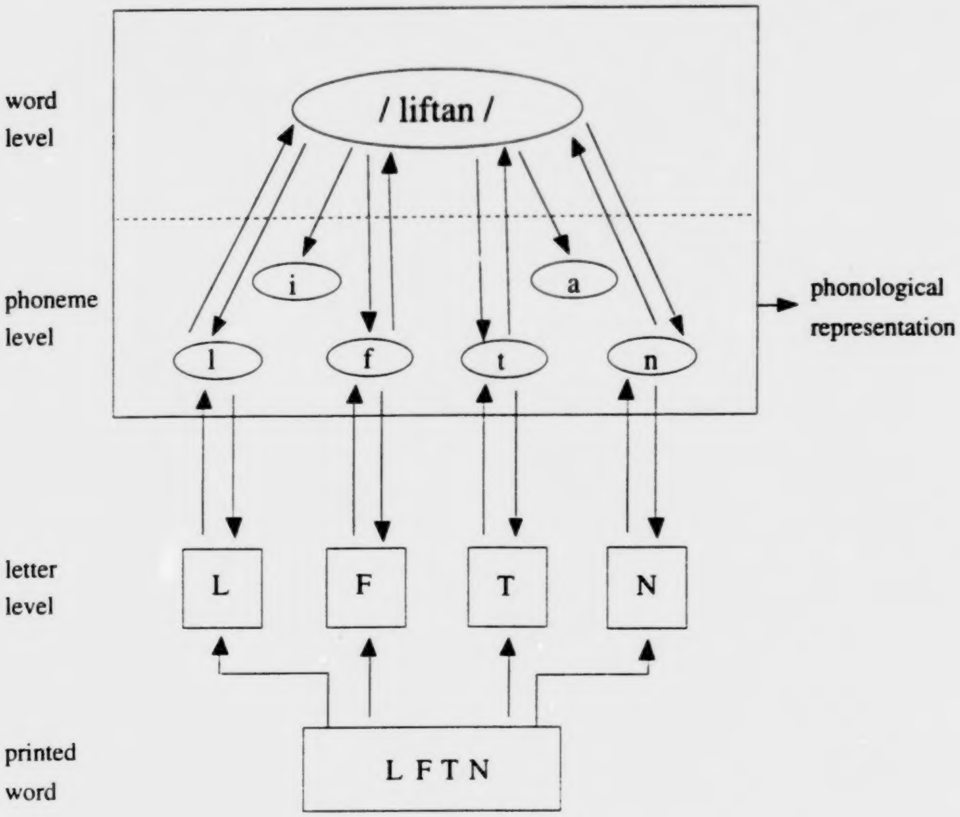


Figure 2. **A model for computing phonology of unpointed Hebrew words.**

The consonants are computed first in a process that is fast and automatic, whereas the vowels, which are the main source of phonological ambiguity, are computed in a subsequent cycle which is less automatic and involves attention-demanding processing. Additional empirical support for a computational process in English has been recently provided by Treiman, Mullenix, and Bijeliac-Babic (1993). Similar to the DF manipulation of the present study, Treiman and her colleagues mapped the spelling-to-sound relations of all CVC words in the English dictionary, and assigned a pronunciation consistency score to the CV or VC subword units. A regression analysis of naming latencies revealed that the consistency of VC subword units had a significant contribution to the prediction of performance in word pronunciation. These results suggest that even for three-letter frequent English words, phonology is assembled rather than addressed as a unit from the lexicon.

Thus, a strong phonological model that accounts for naming in English will regard the initial phase of phonological computation as a conversion of letters into phonemes (unambiguous letters first), by using prelexical conversion rules. This initial phase can only provide the reader with a poor phonological representation, given the depth of the English orthography. This representation is shaped through lexical knowledge to allow a correct pronunciation. Such a model could, in principle, have a similar structure to the model offered by Seidenberg and McClelland (1989), with the difference that lexical information concerning the specific word pronunciation does play an indispensable role in pronunciation.

The mandatory interaction between assembled and lexical phonology has been argued in length by Turvey and his colleagues (e.g., Carello et al., 1992), to account for naming in shallower orthographies like Serbo-Croatian. Several studies in Serbo-Croatian have shown that lexical involvement is apparent even in an extremely shallow orthography. The present study offers a point of reference in the opposite side of the orthographic depth continuum, suggesting that prelexical computation occurs even in the deepest orthographies. By this view, phonology is always assembled and always lexically shaped, but not holistically "addressed".

Admittedly, the weak phonological hypothesis or even the visual hypothesis could, in principle, accommodate the effect of DF on naming, but not without a considerable cost. One could argue, for example, that phonology is retrieved holistically, but the mapping of whole-word orthographic units into whole-word phonological units becomes slower with increasing numbers of missing vowels. In other words, increased numbers of missing vowels in the orthographic representation (high-DF words) could lead to increased difficulty in making contact with the whole-word phonological units in the lexicon. However, this account would deprive "addressed phonology" and "visual access" of their major appeal in reading theory, which is to bypass the many inconsistencies in mapping graphemes into phonemes in deep orthographies. Moreover, by this view, phonology is perhaps "addressed", but in a manner that mimics a piecemeal assembly process. Addressed phonology would consequently retain nothing but its label, losing its theoretical significance.

Another possible interpretation of the DF effect could suggest that for some words phonology was entirely addressed whereas for some words it was assembled. Consequently, the overall DF effect obtained in the present study emerged merely from the subset of words for which prelexical computation was not bypassed by the fast lexical routine. This possibility is not well supported by our results. First, if the DF effect was restricted to a subset of words (presumably very low-frequency, for which addressed phonology has no advantage over assembled phonology) it would not be reliable in the item analysis. Moreover, a clear interaction of DF and frequency would have emerged, in this case, especially in the item analysis. The results of Experiment 1 show an opposite pattern. The main effect of DF was reliable by items whereas the interaction was not. Thus, the present study provides strong support for a model of naming that assumes a mandatory prelexical computation of phonology and a parallel lexical shaping of these computed representations.

The methodology employed in the present study offers a new approach to examine the assembly process, mainly to examine the *size* of the computed units. In principle, such methodology could be implemented in English if an ambiguity score could be computed for each letter slot, and if consequently a DF score could be assigned to each word. This might entail complex computations (But see Treiman, 1993). However the results of the present study suggest that even coarse grained measurements of levels of ambiguity can predict effects on naming latencies.

REFERENCES

Baluch, B., & Besner, D. (1991). Strategic use of lexical and nonlexical routines in visual word recognition: Evidence from oral reading in Persian. *Journal of Experimental Psychology: Learning, Memory, and Cognition. 17*, 644-652.

Bentin, S., & Frost, R. (1987). Processing lexical ambiguity and visual word recognition in a deep orthography. *Memory & Cognition, 15*, 13-23.

Bentin, S., & Frost, R. (in press). Morphological factors in visual word identification in Hebrew. In L. Feldman (Ed.), *Morphological aspects of language processing*. Hillsdale, NJ., Erlbaum.

Berent, I., & Perfetti, C. A. (1993). Roses are reezes: Toward a nonlinear model of phonological assembly in reading. Paper presented at the 34th Annual Meeting of the Psychonomic Society, Washington DC.

Besner, D., & Smith, M. C. (1992). Basic processes in reading: is the orthographic depth hypothesis sinking? In R. Frost & L. Katz (Eds.), *Orthography, phonology, morphology, and meaning* (pp. 45-66). Advances in Psychology, North-Holland: Elsevier.

Carello, C., Turvey, M. T., & Lukatela, G. (1992). Can theories of word recognition remain stubbornly nonphonological? In R. Frost & L. Katz (Eds.), *Orthography, phonology, morphology, and meaning* (pp. 211-226). Advances in Psychology, North-Holland: Elsevier.

Carello, C., Turvey, M. T., & Lukatela, G. (in press). Lexical involvement in naming does not contravene prelexical phonology: A reply to Sebastián-Gallés (1991). *Journal of Experimental Psychology: Human Perception and Performance.*

DeFrancis, J. (1989). *Visible speech: The diverse oneness of writing systems.* Honolulu: University of Hawaii Press.

Feldman, L. B., & Turvey, M. T. (1983). Word recognition in Serbo-Croatian is phonologically analytic. *Journal of Experimental Psychology: Human Perception and Performance, 9*, 228-298.

Frost, R. (1994). Prelexical and postlexical strategies in reading: Evidence from a deep and a shallow orthography. *Journal of Experimental Psychology: Learning, Memory, and Cognition. 20*, 1-14.

Frost, R. (1992). Orthography and Phonology: The psychological reality of orthographic depth. In M. Noonan, P. Downing, & S. Lima (Eds.), *The linguistics of literacy* (pp. 255-274). Amsterdam: John Benjamins.

Frost, R., Katz, L., & Bentin, S. (1987). Strategies for visual word recognition and orthographical depth: A multilingual comparison. *Journal of Experimental Psychology: Human Perception and Performance, 13*, 104-115.

Frost, R., & Katz, L. (1989). Orthographic depth and the interaction of visual and auditory processing in word recognition. *Memory & Cognition, 17*, 302-311.

Frost, R., & Bentin, S. (1992a). Processing phonological and semantic ambiguity: Evidence from semantic priming at different SOAs. *Journal of Experimental Psychology: Learning, Memory, and Cognition. 18,* 58-68.

Frost, R., & Bentin, S. (1992b). Reading consonants and guessing vowels: Visual word recognition in Hebrew orthography. In R. Frost & L. Katz (Eds.) *Orthography, phonology, morphology, and meaning* (pp. 27-44). Advances in Psychology, Elsevier, North-Holland.

Frost, R., & Kampf, M. (1993). Phonetic recoding of phonologically ambiguous printed words. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 1-11.

Katz, L., & Feldman L. B. (1981). Linguistic coding in word recognition. In A. M. Lesgold & C. A. Perfetti (Eds.), *Interactive processes in reading.* (pp. 85-105). Hillsdale, NJ: Erlbaum.

Katz, L., & Feldman L. B. (1983). Relation between pronunciation and recognition of printed words in deep and shallow orthographies. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9,* 157-166.

Katz, L., & Frost, R. (1992). Reading in different orthographies: The orthographic depth hypothesis. In R. Frost & L. Katz (Eds.), *Orthography, phonology, morphology, and meaning* (pp. 67-84). Advances in Psychology, North-Holland: Elsevier.

Koriat, A. (1984). Reading without vowels: Lexical access in Hebrew. In H. Bouma & D. G. Bouwhuis (Eds.), *Attention & Performance X.* Hillsdale. NJ: Erlbaum.

Koriat, A. (1985). Lexical access for low and high frequency words in Hebrew. *Memory & Cognition, 13,* 37-44.

Lukatela, G., Feldman, L. B., Turvey, M. T., Carello, C., & Katz, L. (1989).Context effects in bi-alphabetical word perception. *Journal of Memory and Language, 28,* 214-236.

Lukatela, G., & Turvey, M. T. (1990). Automatic and prelexical computation of phonology in visual word identification. *European Journal of Cognitive Psychology, 2,* 325-343.

Mattingly, I. G. (1992). Linguistic awareness and orthographic form. In R. Frost & L. Katz (Eds.), *Orthography, phonology, morphology, and meaning* (pp. 11-26). Advances in Psychology, North-Holland: Elsevier.

Paap, K. R., Noel, R. W., & Johansen, L. S. (1992). Dual-route models of print to sound: Red herrings and real horses. In R. Frost & L. Katz (Eds.), *Orthography, phonology, morphology, and meaning* (pp. 293-318). Advances in Psychology, North-Holland: Elsevier.

Perfetti, C. A., Bell, L. C., & Delaney, S. M. (1988). Automatic (prelexical) phonetic activation in silent word reading: Evidence from backward masking. *Journal of Memory and Language, 27,* 59-70.

Perfetti, C. A., Zhang, S., & Berent, I. (1992). Reading in English and Chinese: Evidence for a universal phonological principle. In R. Frost & L. Katz (Eds.), *Orthography, phonology, morphology, and meaning* (pp. 227-248). Advances in Psychology, North-Holland: Elsevier.

Seidenberg, M. S. (1985). The time course of phonological code activation in two writing systems. *Cognition, 19,* 1-30.

Seidenberg, M. S. (1992). Beyond orthographic depth in reading: Equitable division of labor. In R. Frost & L. Katz (Eds.) *Orthography, phonology, morphology, and meaning* (pp. 85-118). Advances in Psychology, North-Holland: Elsevier.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed developmental model of word recognition and naming. *Psychological Review, 96,* 523-568.

Shimron, J. (1993). The role of vowels in reading: A review of studies of English and Hebrew. *Psychological Bulletin, 114,* 52-67.

Tabossi, P., & Laghi, L. (1992). Semantic priming in the pronunciation of words in two writing systems: Italian and English. *Memory & Cognition, 20,* 303-313.

Treiman, R., Mullenix, J. W., & Bijeljac-Babic, R. (1993). Spelling-sounds relations in English and their effects on reading. Paper presented at the 34th Annual Meeting of the Psychonomic Society, Washington DC.

Turvey, M. T., Feldman. L. B., & Lukatela, G. (1984). The Serbo-Croatian orthography constrains the reader to a phonologically analytic strategy. In L. Henderson (Ed.), *Orthographies and reading: perspectives from cognitive psychology, neuropsychology, and linguistics* (pp. 81-89). Hillsdale, NJ: Lawrence Erlbaum Associates.

Van Orden, G. C. (1987). A ROWS is a ROSE: Spelling, sound and reading. *Memory & Cognition, 15,* 181-198.

Van Orden, G. C., Johnston, J. C., and Halle, B. L. (1988). Word identification in reading proceeds from spelling to sound to meaning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 371-386.

Van Orden, G. C., Pennington, B.F., & Stone, G. O. (1990). Word identification in reading and the promise of subsymbolic psycholinguistics. *Psychological Review, 97,* 488-522.

FOOTNOTES

[*] *Journal of Experimental Psychology: Learning, Memory, and Cognition,* in press.

[†] Also Department of Psychology, The Hebrew University

[1] The origin of vowel letters reflects an ancient distinction between different forms of vowels that differ mainly in their duration, long vowels being represented by letters. This distinction, however, has no true phonetic reality in modern Hebrew. There are specific grammatical rules that determine when a long or a short vowel should be employed, and consequently these rules specify whether the vowel should be printed with a letter or not. However, because the different printed forms of those vowels do not reflect a phonetic distinction in the spoken language, these rules are often not known to the adult writer and the inclusion of vowel letters is sometimes optional. Consequently, many words may appear with or without the vowel letters in different texts or within the same text (see Shimron, 1993, for a discussion).

[2] The phonologic structure of an unpointed printed word containing vowel letters can be assembled *almost* as easily as in pointed print because the vowel letters still contain some ambiguity. First, the same letter represents both /o/ and /u/. Second, in a few cases the vowel letters can be read as consonants as well; the letter "י" can represent the vowel /i/ but also the consonant /j/, whereas the letter "ו" can represent the vowels /o/ and /u/, but also the consonant /v/. This additional source of ambiguity, however, is limited because these letters are usually doubled to convey the consonant reading.

[3] Given the variance of the frequency ratings (sd = 1.3) these small differences in word frequency were not reliable (F(1,254) <1.0).

[4] Low-DF words have fewer phonemes than high-DF words when the number of letters is kept constant because some of the letters of low-DF words are vowels whereas most letters of high-DF words are consonants. Since each of the consonants can take a vowel, the overall number of phonemes contained in high-DF words is larger.

Haskins Laboratories Status Report on Speech Research
1993, SR-115/116, 127-144

The Tritone Paradox and the Pitch Range of the Speaking Voice: A Dubious Connection*

Bruno H. Repp

Deutsch and coworkers (*Music Perception*, 1990, 7, 371-384; 1991, 8, 335-347) have proposed that individual differences in the perception of the "tritone paradox" derive from listeners' reference to a mental pitch template, acquired through experience with the pitch range of their own voice, as well as with the voice ranges typical of their language community. These authors have reported a correspondence between perceptual results and the upper limit of the individual voice range for a small group of selected subjects, as well as a striking difference in tritone perception between American and British listeners. The present study compared groups of Dutch, British, and American listeners on two tritone tests and also collected voice pitch data for the first two groups in a reading task. There was no within-group correlation of perceptual results with individual differences in voice range. Differences in tritone perception as a function of stimulus characteristics (spectral envelope) were much larger than reported by Deutsch, which casts doubt on the notion of stable individual pitch templates. A significant difference between British and American listeners, with the Dutch group in between, was found in one of the two tritone tests but not in the other. While the origin of this difference remains unclear, it seems unlikely that it has anything to do with regional differences in voice pitch range.

INTRODUCTION

The purpose of the present study was to attempt to replicate and extend the startling and potentially important findings of Deutsch, North, and Ray (1990) and Deutsch (1991) concerning a possible connection between the perception of complex tones and the fundamental frequency range of the speaking voice. The basic theoretical claim is that individuals acquire a stable pitch template from exposure to their own voice and other voices in their language community, and that this template determines the perception of relative pitch height in the task that gives rise to the "tritone paradox."

This experimental paradigm employs complex tones of the kind devised by Shepard (1964) to demonstrate the independence of pitch height and pitch quality (chroma). They are composed of octave-spaced partials whose relative amplitudes are determined by a fixed spectral envelope. Shepard showed that, if the frequencies of all partials are increased in small steps of, say, one semitone (st), listeners perceive successive tones that increase in pitch. These increases continue to be heard indefinitely even though after twelve steps a tone identical with the starting tone is reached, so that the tones keep going around the "pitch (chroma) circle" without ever increasing in pitch height in any objective sense. When pairs of these tones are formed, the resulting musical interval is perceived as rising or falling according

---

This research was conducted in spring of 1993 while the author spent three months as a research fellow at the Institute for Perception Research (IPO) in Eindhoven, The Netherlands. The support of the Technical University Eindhoven and the hospitality of IPO during that period are gratefully acknowledged. Thanks are also due to Adrian Houtsma for providing the software for stimulus generation and the data in Appendix B2, to Bob Crowder for suggesting the method of constructing the context-balanced stimulus sequences, to Rob Meerding for translating the reading materials into Dutch, to Roel Smits for providing the Sennheiser earphones, to Chris Darwin for making it possible for the author to run subjects at the University of Sussex, to Pennie Smith for scheduling those subjects, to Chris Plack and Twan Aarts for technical help, and to all the colleagues at IPO and Haskins Laboratories who served as unpaid volunteer subjects and often provided useful comments. Thanks are further due to René Collier, Bob Crowder, Diana Deutsch, Bill Hartmann, Adrian Houtsma, Brian Moore, Richard Parncutt, Ani Patel, Jacques Terken, and Dix Ward for helpful comments on earlier versions of the manuscript, to Richard Parncutt for additional extensive discussions of this research, and to Bill Hartmann for diplomacy and advice.

to a principle of pitch proximity; thus, for example, the interval C-E is usually heard as a rising major third, not as a falling minor sixth, whereas the interval C-A is heard as a falling minor third, not as a rising major sixth. The ambiguous interval of a half-octave or tritone, such as C-F#, is sometimes heard as rising, sometimes as falling. Shepard pointed out that this ambiguity is rarely perceived as such on any given trial, and he referred to reversible figures such as the Necker cube as a visual analogy.[1]

Deutsch's (1986) tritone test consists of a random sequence of such tone pairs, all of which form tritone intervals but start on any of the 12 semitone steps within one octave.[2] Her novel finding, further documented in several subsequent studies (Deutsch, 1987, 1991; Deutsch, Kuyper, & Fisher, 1987; Deutsch et al., 1990), was that individual listeners, rather than perceiving all these intervals sometimes as rising and sometimes as falling, perceive some of them consistently as rising and others (viz., their inverses) consistently as falling. A typical response function of a hypothetical subject is shown in Figure 1. The consistently rising or falling intervals may not be the same for different listeners, so that the same interval may be heard as rising by one subject but as falling by another (the "tritone paradox"). These results suggested to Deutsch that individuals refer to a personal pitch template, which may be portrayed as a particular orientation of the pitch circle (see Figure 1), such that some pitch classes (the ones on top of the rotated circle) are subjectively "higher" than others. Deutsch (1987) showed that this finding is not an artifact of using a fixed spectral envelope centered on a particular frequency: Stimuli with envelopes centered on different frequencies are perceived similarly, though some listeners do show small shifts in their response functions. Moreover, Deutsch et al. (1987) found that, in a group of American listeners, the individually "highest" pitch classes were not randomly distributed: They fell most often between B and D#, but almost never between F# and A.

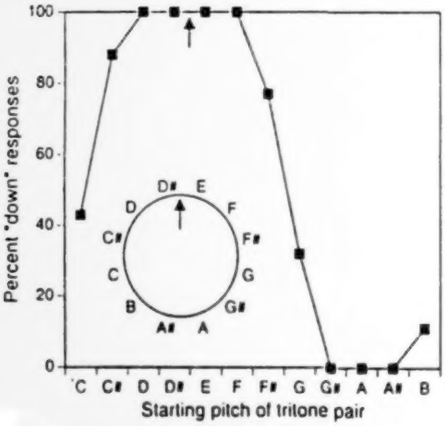


Figure 1. **Response function of a hypothetical subject in the tritone paradigm, and inferred orientation of the subject's pitch circle.**

These interesting observations, which suggested that ordinary listeners possess something like absolute pitch, were followed by two studies in which an explanation of the origin of individual pitch templates was proposed. Deutsch et al. (1990) presented data for a small group of subjects, selected because of their different response functions in the tritone test. Each of these subjects was recorded speaking for about 15 minutes in an interview-like situation, and the speech was analyzed to yield an overall fundamental frequency (F0) distribution. Deutsch et al. then determined for each speaker the octave band that included the largest percentage of the F0 values. For 8 of 9 subjects, the pitch classes delimiting this octave band and the pitch classes perceived as "highest" in the tritone test were within 2 st of each other, a result that significantly deviated from chance.[3] The authors hypothesized that the individual orientation of the pitch circle derives from experience with one's own speaking voice; in particular, that the upper limit of the vocal range defines the pitch classes that are perceived as highest.

An even more remarkable result was reported by Deutsch (1991). In that study, the tritone perception results of two groups of subjects were compared, one from California and the other from southern England (both tested in California under identical conditions). The distributions of the individual pitch circle orientations within each group were strikingly different: Whereas for the American subjects the highest pitch classes were most often between B and D#, for the British subjects they were most often between F# and G#. Thus the two distributions were almost complementary. Deutsch concluded that "perception of music can be strongly influenced by the language spoken by the listener" (p. 345). Although she did not spell out what the relevant regional language characteristic was, the obvious implication seems to be that it lies in the F0 ranges used by American and British speakers.[4] Unfortunately, the study did not contain any speech data.

Differences in intonation between British and American English have been described in some publications, but plots of long-term F0 distributions of the kind examined by Deutsch et al. (1990) are virtually absent from the phonetic literature. It is known that British English makes use of relatively large pitch excursions compared to languages such as Dutch and German (De Pijper, 1983; Willems, Collier, & 't Hart, 1988; Collier, 1991) and, presumably, American English.[5] The implication of this for the long-term F0 distribution is that it will be wider and more skewed towards high frequencies. The average F0 will also be higher. Differences in the average F0 of speakers from different languages or dialect groups have been reported in several studies, though none included British English. However, they suggest that speakers of General American English have relatively low average F0 values, compared to speakers of Southern U.S. dialects (Hanley, 1951), Spanish (Hanley, Snidecor, & Ringel, 1966; Hanley & Snidecor, 1967), and Japanese (Hanley et al., 1966; Hanley & Snidecor, 1967; Yamazawa & Hollien, 1992).

Although a difference in average F0 and/or F0 range between British and American speakers seems plausible, it is also known that F0 characteristics vary widely among adults within any language or dialect community. The most obvious cause of such differences, the sex of the speaker, can perhaps be disregarded in the present context: Women's voices tend to be roughly one octave higher than men's, so that the average male and female F0 ranges are fairly similar in terms of musical pitch classes.[6] Among adults of the same sex, however, there is wide variation in average F0 (see, e.g., Boë & Rakotofiringa, 1975; Hollien & Jackson, 1973; Horii, 1975), due to anatomical (vocal cord length), physiological (age, pathology), psychological (e.g., personality), linguistic, and other factors. Among male speakers, average F0 can differ by as much as an octave; for women, the range of interindividual variation (in st) is somewhat smaller. This variation implies that, even if there is a difference in average F0 or F0 range between two language groups, there will be substantial overlap in the distribution of individual speakers' values. The almost nonoverlapping distributions of the pitch circle orientations of British and American listeners in Deutsch's (1991) study thus seem at variance with the known distributional characteristics of speaking voices.

There are also some potential problems with Deutsch et al.'s (1990) method of capturing intraindividual F0 distributions within an octave band. Although the average width of speakers' F0 ranges tends to be close to an octave (Hanley, 1951; Hudson & Holbrook, 1982), individual ranges vary considerably, which is why studies in the speech literature generally use percentiles or standard deviations to characterize F0 range (see, e.g., Jassem, 1971; Jassem & Kudela-Dobrogowska, 1980). The octave band procedure overestimates the bounds of narrow ranges and underestimates those of wide ranges. Furthermore, Deutsch et al. focus on the upper limit of the F0 range as a potential correlate of the pitch classes that are perceived as highest in the tritone test. This choice is problematic for two reasons. First, the upper limit of an individual F0 distribution is not well defined: The distribution has a fairly gradual slope at high values, and the upper limit is likely to be highly situation-dependent. A more stable point of reference is the *lower* limit of the distribution, usually a fairly abrupt cutoff. Many studies of intonation have pointed out that the bottom of a speaker's range is a stable individual characteristic that is usually reached at the end of a complete utterance (Maeda, 1976; Liberman & Pierrehumbert, 1984; 't Hart et al., 1980; Terken, 1993); no such claim has ever been made about the upper limit of the range, to this author's knowledge. Second, it is not clear why the perceptually highest pitch class should correspond to the highest pitch class in a speaker's octave band, because that pitch class also represents the lower limit of the band. If any pitch class within an (inherently circular) octave range is to be considered "highest", it would have to be one that is several semitones *below* the upper limit, so that it is not only relatively high but also sufficiently removed from the pitch classes at the low end of the range. Therefore, the match between tritone perception and speech production found by Deutsch et al. (1990) may not be so close, after all.

The purpose of the present study was, first, to attempt to replicate the within-group findings of Deutsch et al. (1990) with two separate groups of subjects, one Dutch and the other British, using a sentence reading task to estimate the upper and lower limits of speakers' F0 ranges. Second, the between-group differences in tritone perception reported by Deutsch (1991) were re-examined and extended by comparing three groups of listeners: Dutch, British, and American. Initially only Dutch and British subjects (plus a few Americans) were tested, as this study was primarily carried out in Europe; an American group was added after the author's return to the U.S. for comparison on the perceptual test only. Based on what is known about the intonational characteristics of Dutch and British English, it was expected that Dutch speakers would be more like American than like British English speakers in their use of F0, though considerable overlap of F0 ranges between language groups was expected. If language influences tritone perception, as Deutsch (1991) conjectured, then a difference between British subjects on the one hand and Dutch and American subjects on the other hand should emerge in the tritone test. Moreover, the results for British and American subjects should match those of Deutsch, with the perceptually highest pitch classes being between F# and G# for the British and between B and D# for the Americans. Third, two sets of tones with different spectral envelopes (plus a third set for Dutch listeners only) were included to verify the crucial prerequisite that individual pitch circle orientations are stable across changes in stimulus characteristics (Deutsch, 1987). Without such stability, it would not make sense to look for correlations between perceptual results and speech characteristics.

Methods

Stimuli

Using software developed at IPO by W. M. Wagenaars, three sets of 12 complex tones each were synthesized. The first two sets followed the specifications in Deutsch's publications (see, e.g., Deutsch, 1991); these will be called "Deutsch tones" in the following. Each of these tones had six partials spaced at octave intervals.[7] Their frequencies were varied in semitone steps. A different spectral amplitude envelope was used in each set, one centered at 622 Hz (D#$_5$) and the other at 440 Hz (A$_4$). These two envelopes are illustrated by the right-hand functions in Figure 2.[8]

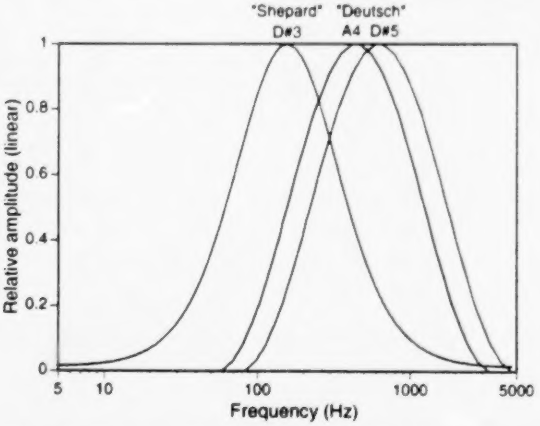


Figure 2. **Fixed spectral envelopes of the Shepard tones (left-hand function) and of two series of Deutsch tones (right-hand functions). Shepard tones had 10 octave-spaced partials under the envelope, Deutsch tones had 6.**

The third set of 12 tones was generated using the specifications in Shepard's (1964) original paper.[9] These "Shepard tones" were included to determine whether they would yield perceptual results equivalent to those obtained with the Deutsch tones. (They were presented only to the Dutch group of subjects.) The Shepard tones had 10 octave-spaced partials (the lowest frequencies actually being below the pitch threshold) and a more peaked spectral envelope centered on 156 Hz (D#$_3$), as shown by the left-hand function in Figure 1. Because of their stronger low-frequency components, the Shepard tones had a fuller, more organ-like timbre than the Deutsch tones. All tones were 500 ms in duration, with 10 ms amplitude ramps at onset and offset. Their overall sound levels were very nearly equal within and across series. All partials started in zero phase. The synthesized waveforms were represented with 16-bit precision at a 10 kHz sampling rate.

Twelve tritone pairs (C-F#, C#-D, etc.) were constructed from each tone series by concatenating the appropriate waveforms without any intervening silence (as in Deutsch's studies). These tritone pairs were then arranged into three tests (Shepard, Deutsch-A, and Deutsch-D#), each comprising 12 repetitions of the 12 pairs of one series. The interpair intervals were 2.5 s, with an extra 2.5 s after each block of 12. Each pair occurred once in each block. Because pilot observations had suggested the existence of strong sequential context effects (see Appendix B-1), the sequence of pairs was such that each pair was preceded once by each other pair, but never by itself.[10] (The initial pair in each block was not scored.) The three 144-item tests used the same context-balanced stimulus sequence, but the order of blocks was different. The test sequences were low-pass filtered at 4.9 kHz and recorded onto digital tape for presentation. Each test was preceded by an ascending ordered sequence of the 12 tones, to mark the beginning and to introduce the subjects to the sound of the tones.[11]

In addition to the listening tests, a set of 10 sentences, both in English and in Dutch translation, was devised for the assessment of F0 characteristics. The sentences were statements of medium length and had relatively de-accented words at the end. They are listed in Appendix A.

Subjects and procedure

Dutch subjects. These were 15 members of the IPO research staff, 8 men and 7 women, all unpaid volunteers and native speakers of Dutch. They listened to all three tests in a quiet classroom using Sennheiser HD 530 II earphones with circumaural cushions. Presentation was binaural at a comfortable intensity (determined with a sound level meter and earphone coupler to be approximately 77 dB SPL). Written instructions were given in English. Some examples, taken at random from within the first test, were played for familiarization before starting. Subjects recorded their judgments of whether the pitch went up or down by writing upward or downward pointing arrows onto an answer sheet; a forced choice had to be made on each trial. There were breaks of a few minutes between tests. The three tests were presented in three different orders to groups of 5 subjects each.

Immediately after the session or at a later time, each subject was asked to read the 10 Dutch sentences in front of a microphone. The sentences were printed on index cards that were randomly shuffled for each subject. The subjects were asked to familiarize themselves with the sentences and then to read them "in a natural and relaxed fashion." The speech was recorded on digital tape for later analysis.

British subjects. They were 10 staff members of the University of Sussex at Brighton, 5 men and 5 women, all natives of southern England who were paid for their participation, and one additional male volunteer (JS1-m, a native of London) who visited IPO and was tested there. The first 10 subjects were tested individually in a sound-isolated booth in the Laboratory of Experimental Psychology at the University of Sussex. The stimulus tape was played back on a portable Sony DAT recorder, and subjects listened binaurally over a pair of the earphones used at IPO, which had been brought along by the author. The voltage at the earphones was calibrated with a volt meter to equal the value measured at IPO (70 mV). Only the two Deutsch tone tests were presented to the British subjects, 6 listening in one order and 5 in the other. Otherwise, the procedure of testing and recording was the same as for the Dutch subjects, except that the sentences were read in English.

American subjects. There were 17 subjects, 8 men and 9 women, all unpaid volunteers. One (GS-m) was a postdoc at the University of Sussex; four others (DK-m, JJ-m, SC-f, TM-f) were visiting researchers at IPO. These 5 subjects were tested under the same conditions as the British and Dutch subjects, respectively, except that they listened only to the two Deutsch tone tests. They were also recorded reading the English sentences. The remaining 12 subjects were 9 members of the research staff at Haskins Laboratories, two additional graduate students, and one recent high school graduate.[12] They were tested individually in a quiet room at Haskins Laboratories using Sennheiser HD 420 SL earphones with on-the-ear cushions. The playback level was similar to (but not calibrated to be identical with) that used in Europe. (See Appendix B-2 concerning effects of playback level.) Only the two Deutsch tone tests were presented, with their order varying across subjects. The ascending tone sequence preceding each test was omitted for these 12 subjects. Also, no speech samples were collected from them, as the IPO software used for the F0 analysis was no longer available.

Data analysis

For each subject and each tritone test, the number of "down" responses to each of the 12 stimulus pairs was tallied (ignoring the first response in each block). Subsequently, the six adjacent stimulus pairs were determined which had the highest number of "down" responses (see Figure 1). Following Deutsch's procedures, the starting pitches of the middle two (i.e., the third and fourth) of these pairs were taken to be the "highest" pitches, regardless of the exact distribution of responses. In most cases, these pairs in fact received the highest number of "down" responses, although adjacent pairs often had equal (maximum) scores. Although the clarity of the "pitch class effect" (i.e., the complementary distribution of "up" and "down" responses across the 12 stimulus pairs) varied, the subjectively highest pitch classes could be determined in this way in all but four instances of apparently random responses.[13]

The recorded speech was digitized at 10 kHz with low-pass filtering at 4.9 kHz, and F0 contours were determined using software developed at IPO which employs the method of subharmonic summation (Hermes, 1988). Two F0 values were measured in each sentence contour: The lowest value near the end (taking care to avoid artifacts due to vocal fry or very low amplitude), and the highest value, which was usually on the first accented syllable. These values could be determined in all but one instance (a missing sentence). In four additional instances, one of the two measurement values was excluded as an extreme outlier compared to the values from the other 9 sentences. One Dutch subject's (JR-m) extremely low F0 trailed off into creak at the ends of sentences, so only the end of regular voicing could be measured. The measured values of each subject were averaged across the 10 sentences, and standard deviations were calculated (in linear Hz).

Results and Discussion

Tritone perception

Individual pitch class effects. The existence of pitch class effects for individual subjects was amply confirmed by the present results. As already mentioned, there were only four instances in which it was impossible to determine a particular orientation of the pitch circle.[14] There was no subject who failed to show a pitch class effect on all two or three tests, though there were a few who showed weak effects throughout. (See also Appendix B-1.) Defining as a "clear" effect any response pattern that showed 0 or 100 percent "down" responses for at least one stimulus pair, 28 of the 45 Dutch (15 subjects x 3 tests), 14 of the 22 British (11 subjects x 2 tests), and 26 of the 34 American (17 subjects x 2 tests) test results fell in that category. Defining as a "strong" effect any response pattern with 6 or more stimulus pairs receiving extreme scores (as in the example in Figure 1), there were 18 Dutch, 6 British, and 16 American cases in that category. The results also show that reliable pitch class effects can be obtained with the original Shepard tones: Of the 15 Dutch subjects, 8 showed a clear pitch class effect in the Deutsch-D# test, 11 in the Deutsch-A test, and 9 in the Shepard test. However, pitch class effects tended to be more pronounced in the Deutsch-A test than in the Deutsch-D# test: In the former, there were 23 strong, 10 moderate (i.e., clear but not strong), and 8 weak effects overall, while in the latter there were 9 strong, 19 moderate, and 14 weak effects.

Between-test differences. To present one extreme but not singular example of individual response functions, Figure 3 shows the results of one Dutch subject (MS-f) who showed clear pitch class effects in all three tests. The results of the three tests were extremely dissimilar, which was quite unexpected given the data reported by Deutsch (1987). Results such as shown in the figure could not possibly reflect a stable individual pitch template that is operative regardless of the spectral characteristics of the tones. Clearly, the spectral envelopes of the tones made a substantial difference.

The tritone test results of all subjects are summarized in Figure 4 in terms of the individually highest pitch classes in each test. It is evident that differences among the three tests were pervasive, systematic, and frequently very large. Of the 40 subjects who yielded at least weak pitch class effects in both Deutsch tone tests, only 3 (all American) showed identical effects, whereas 5 (1 Dutch, 4 American) produced *maximally different* effects (i.e., separated by 6 st).

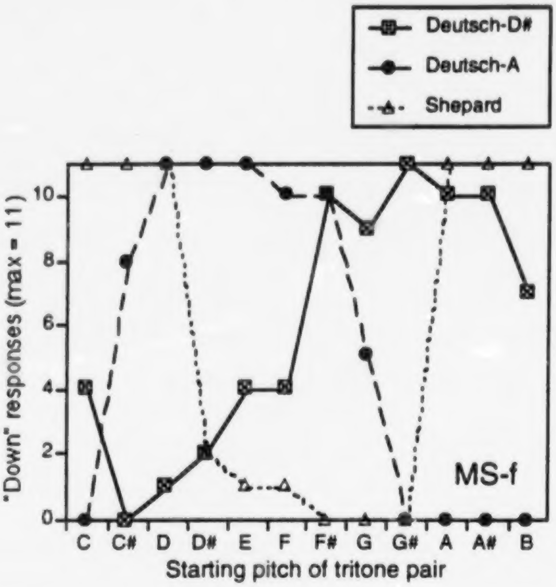


Figure 3. **Response functions of one Dutch subject (MS-f), illustrating large between-test differences.**

Of the remaining 32 subjects, 24 showed a highest pitch class in the Deutsch-D# test that was 1-5 st *higher* than that in the Deutsch-A test; this is significant by a binomial test ($p < .005$) and shows the direction of the difference to be systematic. This pattern was especially characteristic of the Dutch (12 out of 14) and British (8 out of 9) subjects, though not of the Americans (4 out of 9). Breaking down the results in yet another way: Of the 40 subjects, 17 (7 Dutch, 4 British, 6 American) showed similar pitch class effects (0-2 st difference) in the two Deutsch tone tests, 11 (4 Dutch, 5 British, 2 American) showed moderately different effects (3-4 st difference), and 12 (4 Dutch, 8 American) showed highly dissimilar effects (5-6 st difference). Thus, less than half of the subjects showed the pattern Deutsch's reports would have led one to expect.

In addition, it is clear that the Shepard tone test, which was administered only to the Dutch subjects, yielded substantially different results from the two Deutsch tone tests. In only 6 out of 28 possible between-test comparisons were the pitch class effects similar (0-2 st apart), while they were highly dissimilar (5-6 st apart) in 12 comparisons. Only one subject (JT-m) showed similar pitch class effects in all three tests. Overall, these results call into question the notion of an individually stable pitch template and suggest strongly that the pitch class effect depends, at least in part, on spectral stimulus characteristics.

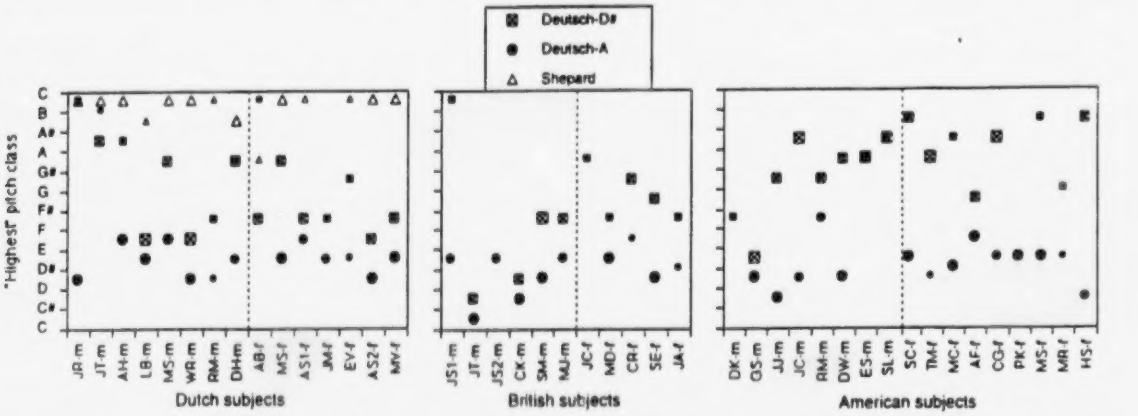


Figure 4. **Summary of tritone test results for 15 Dutch, 11 British, and 17 American subjects. The size of the data points indicates the strength of the effects: strong, moderate, or weak. The subjects within the Dutch and British groups are arranged according to increasing lower limits of their voice range (as in Figure 5). The first three male and first two female American subjects (tested in Europe) are arranged in the same way; the others are ordered arbitrarily. The dotted vertical lines separate male and female subjects.**

One point of concern was that the ascending scale that preceded each test may have biased subjects' responses. Given that preceding context has an effect in these tests (see Appendix B-1), it might be argued that the scale primed listeners to consider certain pitch classes as low and others as high; and since the two Deutsch tests were preceded by different scales, this might have contributed to the large between-test differences in results. However, this "priming hypothesis" can almost certainly be dismissed. First, the Deutsch-D# and Shepard tone tests were preceded by the same scale (D#-D) and so should have yielded similar results, which they did not. Second, the priming hypothesis predicts a particular orientation of the pitch circle: For a preceding D#-D scale, the "highest" pitches should be in the B-C# region, where they were very rarely in the Deutsch-D# test (though the prediction is accurate for the Shepard tone results); and for a preceding A-G# scale, they should be near F-G, which does not fit the data of the Deutsch-A test. The agreement in the Shepard tone test then is more likely a coincidence. Third, it seems highly implausible that listeners would be able to remember the pitch range of the initial scale for very long, given the interference and the context effects caused by the tritone pairs in the test. It seems equally unlikely that a context effect initiated by the scale would be propagated deterministically through the test sequence. Finally, it will be recalled that the precursor scales were omitted in the tests given to the 12 American subjects tested at Haskins Laboratories. Although three of these subjects showed very similar pitch class effects on the two Deutsch tone tests, others showed very large differences. Both patterns of results were also shown by the American subjects tested in Europe (GS-m, JJ-m, SC-f, TM-f), who did hear the precursor scales. Therefore, the scales probably had little effect, if any, on tritone perception.

Within-test differences. There were some striking individual differences within tests, which is consistent with Deutsch's reports. In two of the tests, however, the majority of the subjects showed very similar results, and only a few listeners deviated from this predominant pattern. Individual differences seemed most constrained with the Shepard tones: 13 of the 15 Dutch subjects had their highest pitch classes between A# and C, and for 11 of them they were B and C. One of the two subjects who were not scored (JM-f) had an ambiguous response function that could be interpreted as a B-C pitch class effect as well. The single outlier (subject AB-f) was within 3 st and represented a weak effect by the definition given earlier. Thus, there was actually very little individual variability in the Shepard tone test.

Similarly, the Deutsch-A test showed considerable consistency among subjects. Of the 41 subjects who showed a pitch class effect, 35 had their highest pitch classes between C# and F, 26 between D and E. Two additional subjects were just 1 st away. Both of the two Dutch outliers (JT-m, AB-f) showed very weak effects. This leaves only two truly deviant subjects, the Americans ES-m and SL-m, both of whom showed strong pitch class effects that were about 6 st away from the dominant region—i.e., reversed effects with respect to most other subjects.

Only the Deutsch-D# test showed individual variability of the magnitude Deutsch's results (especially Deutsch et al., 1987) would have led one to expect. Most highest pitch classes (35 out of 42) fell within the half-octave region between E and A#. Five other subjects were within 1 st, and the two clearest outliers (Dutch subject JR-m and British subject JS1-m) again represented weak and hence rather unreliable effects.

The unexpectedly restricted within-test variation on the Deutsch-A and Shepard tone tests obviously provides a very poor basis for investigating correlations with voice range. As will be shown below, subjects varied greatly in their voice characteristics; yet they perceived the tritone pairs in these two tests very similarly. Only the Deutsch-D# test results gave any hope of finding a correlation.

Between-group differences. It is evident from Figure 4 that the striking difference found by Deutsch (1991) between British and American listeners was not replicated in the Deutsch-A test. In fact, there was high agreement among all three subject groups on this test (notwithstanding the presence of two outliers each in the Dutch and American groups). In the Deutsch-D# test, where there was more variability, the results of the three subject groups also overlapped substantially, but there was a definite tendency for American subjects' results to be higher up on the pitch scale than British subjects. This tendency was weakened by the very "high" pitch class effect of British subject JS1-m. However, since the starting pitch class (C) of the circular octave range displayed in Figure 4 is arbitrary, JS1-m could also be imagined just below the abscissa of the graph, in order to maximize the group difference. For the statistical comparison, therefore, the individually highest pitch classes were expressed numerically (C=0, ... B=11), but the results for British subject JS1-m as well as for Dutch subject JR-m were coded as -0.5. With the cards stacked in this way, there was indeed a significant difference among the three subject groups in a one-way ANOVA [$F(2,39) = 5.47$, $p < .009$], and subsequent pairwise comparisons showed this effect to be due mainly to the difference between British and American subjects [$F(1,25) = 10.81$, $p < .004$], with the difference between Dutch and British subjects being nonsignificant [$F(1,23) = 1.56$], and that between Dutch and American subjects being marginally significant [$F(1,30) = 4.61$, $p < .05$]. By contrast, for the similarly coded results on the Deutsch-A test, there was no effect of language group [$F(2,38) = 1.43$].

The average numerically coded highest pitch classes on the Deutsch-D# test were 6.2 (i.e., just above F#; s.d. = 2.6 st) for the Dutch subjects, 4.8 (just below F; s.d. = 2.8 st) for the British subjects, and 8.0 (G#; s.d. = 2.2 st) for the Americans. Thus the average difference between the British and American groups was 3.2 st, which is not inconsistent with the results of Deutsch (1991), though the difference appeared to be larger there. Her British subjects had their highest pitch classes predominantly between F# and G#, which is consistent with the Deutsch-D# test results, where 7 out of 10 British subjects fell between F and G#. However, Deutsch's American subjects fell mostly in the region between B and D#, where none of the present American subjects could be found. Of the 17 Americans, 14 had their highest pitch classes between F# and B in the Deutsch-D# test. The Deutsch-A test results, on the other hand, clash with Deutsch's results for British subjects, while being more compatible for Americans.

Speech data

Voice ranges. Figure 5 shows the individual F0 ranges for the three subject groups, arranged in terms of increasing lower limits within each group. (Speech data were available from only 5 American subjects, those tested in Europe.) The ranges around the estimates of the lowest and highest F0 values for each subject represent one standard deviation above and below the mean computed across the 10 sentences; the distance between the two estimates is the individual vocal range. As expected, most speakers showed quite consistent values for the lower limits of their ranges. The estimates of the upper limits were more variable, as they depended more on the linguistic structure of the sentences.

A considerable diversity of vocal ranges was represented, especially among the male speakers. Quite a number of speakers had ranges of roughly one octave. However, there were some speakers with exceptionally wide or narrow ranges, for whom the octave band approach of Deutsch et al. (1990) would not be appropriate. As expected, there was extensive overlap between both the lower and the upper limits of the voice ranges of Dutch and British (and the few American) subjects. Although it had been predicted that British speakers would exhibit higher upper limits and wider individual ranges than Dutch and American speakers, there was no indication in the data of such a difference. It seems possible that observations in the linguistic and phonetic literature on the wide pitch excursions of British English apply only to "received pronunciation", a traditional upper-class style of speaking that was not prevalent among the younger generation to which all but one subject belonged.

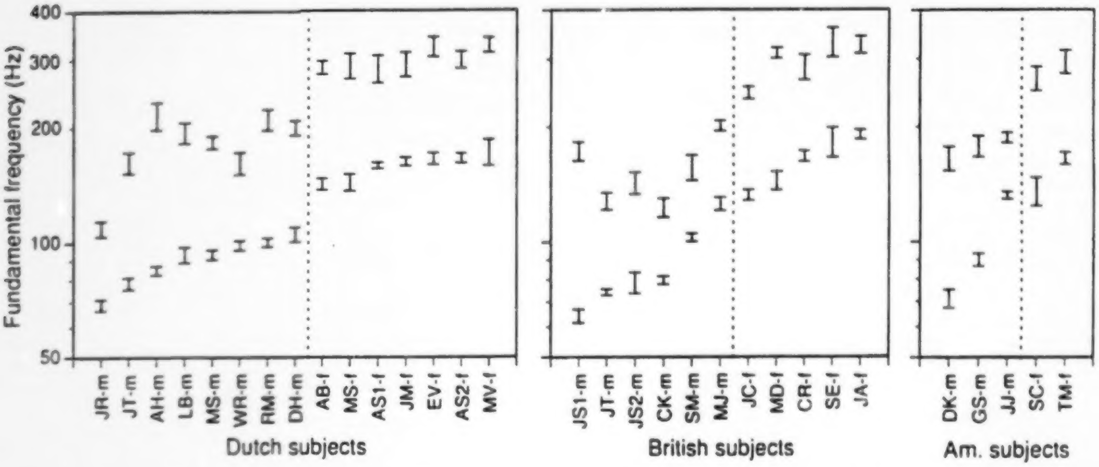


Figure 5. **Estimated values (ranges of one standard deviation above and below the mean) of the lower and upper limits of subjects' speaking ranges. The dotted vertical lines separate male and female speakers.**

The single middle-aged subject, JS1-m, was the only British speaker who showed an exceptionally wide vocal range.

Two-way ANOVAs were conducted on upper limits, lower limits, and their difference; the factors were language (Dutch vs. English) and speaker sex. The female F0 values were first divided by 2, thus lowering them by one octave. There were no significant effects in any of these analyses.

Relationship between voice range and tritone perception. The cross-language comparison of vocal ranges offers no clues as to what aspect of the speaking voice might account for the differences between British and American listeners in the Deutsch-D# tritone perception test. Admittedly, this aspect of the study is hampered by the absence of speech data for the majority of the American subjects. Note, however, that the perceptually highest pitch classes were *higher* for American than for British subjects (Figure 4), whereas their voice ranges were expected to be lower. This makes it seem very unlikely that the group differences in tritone perception have any F0 correlate.

Because of their variability, the F0 data provide a good basis for exploring correlations with individual subjects' perceptual results. Unfortunately, as was mentioned earlier, the results for two of the tritone tests (Deutsch-A and Shepard) do not offer enough individual variability for that purpose. Therefore, the analysis focuses on the Deutsch-D# test results.

Deutsch et al. (1990) hypothesized that the upper limit of the vocal range—and, by implication via their octave-band criterion, the lower limit as well—should match the pitch classes on top of the individual pitch circle. This prediction is tested in Figure 6, where the highest pitch classes obtained in the Deutsch-D# test are plotted as a function of the individual lower and upper voice limits, respectively, for all subjects that provided speech data. The solid diagonal line represents identity, and the parallel dotted lines delimit a range of plus/minus 6 st. (These lines coincide on the cylindrical surface that is flattened out in the figure.) It is evident that there was no close correspondence of the perceptual data with either end of the voice range, although in each case there was a slight trend: 18 out of 30 data points fell within 3 st of the lower limit of the voice range (n.s.), and 20 out of 30 fell within 3 st of the upper limit ($p = .05$).

There is a problem with these comparisons, however: As pointed out in the Introduction, it is not clear why either the highest or the lowest pitch class in the voice range should correspond to the highest perceived pitch class. When the voice range equals or exceeds an octave, the highest pitch class(es) is (are) simultaneously the lowest pitch class(es). The truly highest pitch classes in a circular range are those that have no representation at the low end. Allowing for some "smoothing" or uncertainty at each end of the range to avoid abrupt discontinuities, the truly highest pitch class should be about 9 st above the lower limit. This more sensible pitch class reference is indicated by the dashed line in Figure 6a. It does not correlate with the perceptual data, as only 16 out of 30 data points fall within 3 st of it.

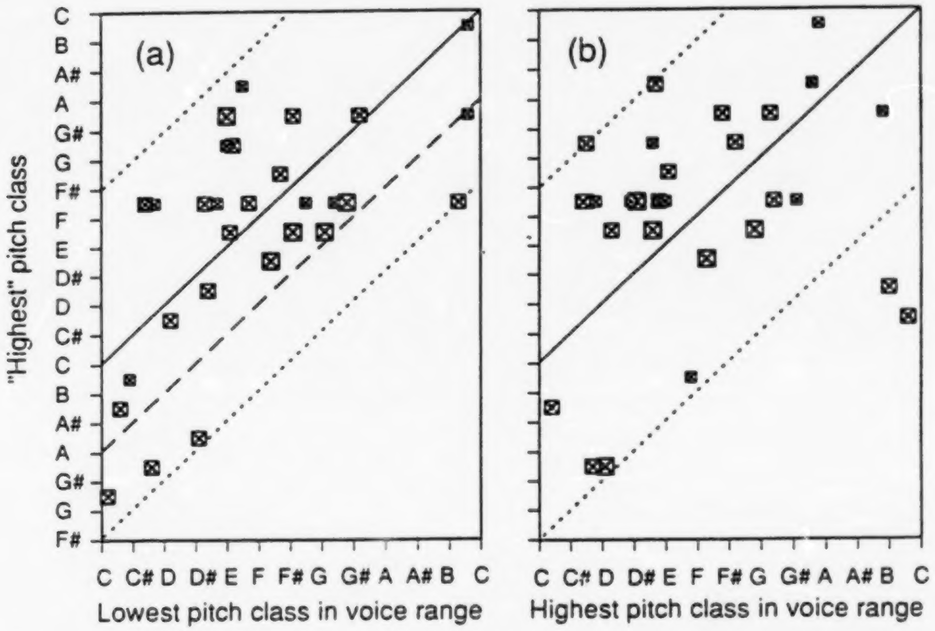


Figure 6. **Mean values of subjects' (a) lowest and (b) highest voice pitches plotted against the perceptually highest pitch classes in the Deutsch-D# tritone test. Results for 15 Dutch, 10 British, and 5 American subjects combined. The size of the symbols indicates the strength of the pitch class effect (as in Figure 4).The dotted lines delimit the octave band around the line of equality (solid diagonal). The dashed line indicates the pitch classes 9 st above (3 st below) the lower limit.**

GENERAL DISCUSSION

Deutsch's findings on the tritone paradox are extremely interesting because they seem to suggest that ordinary listeners without absolute-pitch capacities nevertheless have a stable pitch reference in their heads. Three findings seemed well established before the present study was conducted: (1) Individual listeners show pitch class effects in the tritone task, (2) individual pitch class effects are essentially stable across tests using tones with different spectral envelopes, and (3) there is large individual variability in pitch class effects.

The present study confirms the first result but unexpectedly challenges the second one and, in part, the third. In the three different tests employed here, and in the two Deutsch tone tests in particular, widely divergent results were obtained for many subjects. Moreover, these differences followed a fairly systematic pattern. These results not only suggest that spectral envelope characteristics play an important role in the tritone perception task, but that the assumption of an individually stable pitch template may be incorrect.

The reason for this apparent discrepancy from Deutsch's findings is not known at present. Deutsch (1987) reported data for only four subjects. These subjects listened to 12 different stimulus sets differing in the center frequencies of their spectral envelopes, two of which (D#5 and A4) matched the present Deutsch-D# and Deutsch-A sets. For these two tests, the subjects' "highest" pitch classes did not differ by more than 1 st, and response functions were generally similar across all 12 tests. Deutsch et al. (1987: Figure 7) display results for a single subject whose particularly pronounced pitch class effects were virtually identical in four different tests. In her later studies, Deutsch averaged over the results of several tests, which suggests that the pitch class effect varied only slightly and idiosyncratically as a function of spectral envelope. Whether that was in fact the case in all instances has not been documented. In the present study, however, it would make little sense to average over the Deutsch-D# and Deutsch-A test results; the discrepancies are much too large.

Another unexpected finding was the restricted individual variability in the Shepard and Deutsch-

A tests. It appears that tones with certain spectral characteristics are perceived similarly by most or all listeners, which is contrary to the hypothesis that tritone perception reflects diverse individual pitch templates. It could be, however, that spectral and individual determinants of tritone perception are in competition, with the former gaining the upper hand in certain situations. This does not increase one's confidence in the Deutsch-D# test results as pure measures of individual pitch templates; on the contrary, spectral stimulus properties probably had an influence on subjects' percepts in that test also.

Deutsch's more recent findings regarding the connection between tritone perception and speech characteristics seemed not as well established and open to criticism even before the present study was conducted (see the Introduction). Her report of a correspondence between the upper limits of subjects' voice ranges and their hypothetical pitch templates (Deutsch et al., 1990) was based on a small group of selected subjects and on a questionable method of estimating the relevant pitch class in the voice range. Although her selection of subjects with very different pitch class effects is methodologically defensible, it does neglect subjects who have similar pitch class effects but different voice ranges—cases that are inconsistent with the hypothesis being tested. Although her most recent study (Deutsch, 1991) presented a striking difference between British and American subjects in tritone perception, it provided neither speech data nor an explicit hypothesis about the way in which tritone perception might depend on "language".

The present study sought to fill some of these gaps but was hampered by the unexpected instability of the pitch class effect across different tests, as well as by the lack of sufficient individual variability within two of the tests. Only the Deutsch-D# test yielded results that could be used to address the issues raised by Deutsch, which naturally weakens the conclusions. There is no reason why the pitch class effect exhibited on that particular test should reflect *the* pitch template of any given listener. Given that caveat, it is perhaps not surprising that there was no clear correspondence between the perceptual results and estimates of the highest pitch classes in subjects' voice ranges.

It could be objected that the current voice range estimates were inaccurate, either because of the small number of speech samples or because of the artificiality of the reading situation. Both objections are probably valid with regard to the upper limit of the range, which is not well defined and changes with speakers' emotional state and other factors. However, as was argued in the Introduction, the lower limit of the range is a more stable reference point, and that parameter was probably estimated with sufficient accuracy in the present study. It was also argued here that the highest pitch class in a subject's range should be defined relative to the lower, not the upper limit. As to the situation dependence of vocal range, studies in the speech literature have consistently reported somewhat higher average F0 values for reading than for impromptu speaking (e.g., Hanley, 1951; Hanley et al., 1966; Mysak, 1959; Snidecor, 1943), but the difference is negligible at the lower end of the vocal range (Hollien & Jackson, 1973; Hudson & Holbrook, 1982).

While the present results conflict with Deutsch's findings in several ways, there seems to be agreement with regard to a difference in tritone perception between British and American listeners. This agreement is only superficial, however. First, it derives from the Deutsch-D# test alone; on the Deutsch-A test, some other factor seemed to constrain perception similarly in both groups. Second, although the predominant "highest" pitch classes in the Deutsch-D# test agreed with those found by Deutsch (1991) for British subjects, there was no such match for Americans. It could be argued that this disagreement is due to the fact that Deutsch's subjects were all from California, whereas the present Americans were a heterogeneous group from various parts of the country (including California).[15] Ragozzine and Deutsch (1993) have recently reported some evidence for regional differences in tritone perception within the United States (in fact, within Youngstown, Ohio), suggesting that the regional origin of a subject's parents needs to be taken into account in interpreting the results. Be that as it may, it is far from clear how such regional differences are to be explained. Corroborative data on corresponding differences in regional speech characteristics have not been presented so far.

Third, the difference between British and American subjects in tritone perception remains puzzling because of the apparent absence of any corresponding difference in voice ranges between the groups. To be sure, not much can be concluded from the present comparison of 11 British speakers with only 5 Americans (those for whom speech data were available). However, by all accounts, Americans should have *lower* voice ranges than speakers of British English, whereas the Deutsch-D# test results indicated *higher* "highest" pitch classes for Americans. Thus it seems extremely unlikely that voice characteristics could account for this difference.

Fourth, the between-group difference occurred in the absence of any within-group correlation with voice characteristics. If the pitch class effect had anything to do with individual voice range, then within-group and between-group correlations between tritone perception and speech production should go hand in hand. If so, however, it should have been much more difficult to find a significant between-group difference, considering the large within-group variability in voice characteristics (within each gender) which would seem to entail a similarly large variability in tritone perception.

Given the discrepancies between the present findings and Deutsch's results, it is perhaps premature to try to come up with alternative explanations of the pitch class effect. There may be unsuspected methodological factors that explain these discrepancies, and new data may resolve the matter in favor of Deutsch's model. However, alternative models that treat pitch as a linear rather than as a circular dimension might be considered.

Terhardt (1991) has discussed briefly the tritone paradox from the perspective of his well-known virtual-pitch theory (Terhardt, Stoll, & Seewann, 1982a, 1982b). The theory holds that a complex tone evokes a number of competing pitch percepts, some of them spectral (i.e., directly corresponding to partials), others virtual (i.e., not necessarily corresponding to partials that are present, as in the case of a "missing fundamental"). "Analytic" listeners are more prone to pay attention to spectral pitches (see Appendix B-3), whereas "synthetic" listeners pay more attention to virtual pitches. In the case of Shepard (or Deutsch) tones, the dominant virtual pitches coincide with partials (Terhardt, Stoll, Schermbach, & Parncutt, 1986). Their relative dominance is determined by a spectral weighting function that peaks around 700 Hz. This weighting function effectively gets convolved with the actual spectral envelope of the signal, although Terhardt's model incorporates thresholds below which changes in physical amplitude are assumed to have no perceptual consequences. Terhardt et al. (1986) asked listeners to match pure tones with the perceived virtual pitches of tones with octave-spaced partials; the resulting response distribution ranged from about 200 to 1000 Hz, with a peak around 300 Hz.[16] However, the complex tones in that study had a flat spectral envelope. The pronounced amplitude differences in stimuli with bell-shaped envelopes may well have a significant effect on pitch perception. The perceived pitch of such a tone is likely to be a joint function of the spectral envelope and of the listener's pitch weighting function.[17]

Given a smooth probability distribution of candidate virtual pitches, there must be one Shepard or Deutsch tone in a set of 12 whose two most salient virtual pitches (12 st apart) are equally strong, straddling the peak of the function. The tone whose partials are shifted by 6 st will then have a single most prominent virtual pitch near the peak of the function, and the pair formed by these two tones will be maximally ambiguous as to the direction of the pitch change. Pairs of other tones in between will be "unbalanced" in the sense that, in tritone pairs, they are perceived as changing pitch in a certain direction (i.e., their strongest pitches are on opposite sides of the maximum of the probability distribution). The occurrence of a pitch class effect is thus predicted by Terhardt's model (cf. Terhardt, 1991: Figure 5).

It is also easy to incorporate individual differences in the model by postulating individual variability in the spectral weighting function. Large individual differences in the relative perceptual importance of the partials of complex harmonic tones have been reported by Moore, Glasberg, and Peters (1985). The origin of these differences is not yet clear, however. Terhardt (1991) speculates, like Deutsch et al. (1990), that individual differences may derive from differential exposure to voices, both one's own and those of others. Although he is not specific about the relevant voice characteristics, it should be noted that the spectral weighting function in his model peaks around 700 Hz, which is much higher than the F0 of adult human speakers. F0 may thus not be the relevant characteristic; rather, it may be the long-term speech spectrum in the region of the first formant. Very little is known at present about differences in long-term speech spectra among individuals or across languages, and any more specific hypothesis would be pure speculation. However, it should be noted that, once pitch is treated as a linear rather than circular dimension, sex differences enter the picture, due to the spectral consequences of women's smaller vocal tracts and higher voices. In the tritone task, however, sex differences are generally absent. This suggests that experience with the sound of one's own voice is not a likely factor.

Some incidental observations made in the course of the present study may provide clues to the nature of the pitch class effect. Appendix B presents some of them; others have been mentioned in footnotes. The reduced individual differences and stronger pitch class effects in the tests employing lower-pitched stimuli (Shepard and Deutsch-A tests) deserve attention. Most intriguing, perhaps, are some hints in the present data that the pitch class effect can be reversed between and within individuals. Although a direct test of strategy-based reversal within listeners failed (see Appendix B-3), there was one subject who spontaneously reversed his responses during the Deutsch-D# test (see Footnote 13) and another who apparently did so in the last block of the Deutsch-A test. Some of the weak pitch class effects observed may have been due to more frequent reversals during a test, rather than to random responding. (See also Appendix B-1.) In the American group of subjects, there were several subjects who showed identical or very similar results on the two tests, although most other subjects showed nearly opposite pitch class effects. The pitch class effects of the American subjects thus seem to be bimodally distributed across the two tests, with the modes being about 6 st apart. One is led to wonder how stable the direction of these pitch class effects is. If future studies succeeded in demonstrating that they are reversible within the same listeners, the significance of the "highest" pitch class would be eroded; instead, the *most ambiguous* stimulus pairs would emerge as the common "hinges" of two diametrically opposed pitch class effects. This would be damaging to Deutsch's theory of a stable pitch template. However, there is no convincing evidence at present that individual pitch class effects are in fact reversible.

In summary, the present study has attempted to replicate some of Deutsch's findings and has failed in most respects, though a difference in tritone perception between British and American listeners was confirmed. Further research is needed to clarify the conflicting results and to provide new evidence bearing on Deutsch's provocative theory.

REFERENCES

Boe, L., & Rakotofiringa, H. (1975). A statistical analysis of laryngeal frequency: Its relationship to intensity level and duration. *Language and Speech, 18,* 1-13.

Collier, R. (1991). Multi-language intonation synthesis. *Journal of Phonetics, 19,* 61-73.

De Pijper, J. R. (1983). *Modelling British English intonation.* Dordrecht, The Netherlands: Foris.

Deutsch, D. (1986). A musical paradox. *Music Perception, 6,* 115-132.

Deutsch, D. (1987). The tritone paradox: Effects of spectral variables. *Perception & Psychophysics, 41,* 563-575.

Deutsch, D. (1991). The tritone paradox: An influence of language on music perception. *Music Perception, 8,* 335-347.

Deutsch, D. (1992a). The tritone paradox: Implications for the representation and communication of pitch structures. In M. R. Jones & S. Holleran (Eds.), *Cognitive bases of musical communication* (pp. 115-138). Washington, DC: American Psychological Association.

Deutsch, D. (1992b). Paradoxes of musical pitch. *Scientific American, 267(2),* 88-95.

Deutsch, D. (1992c). Some new pitch paradoxes and their implications. *Philosophical Transactions of the Royal Society of London, Series B, 336,* 391-397.

Deutsch, D., Kuyper, W. L., & Fisher, Y. (1987). The tritone paradox: Its presence and form of distribution in a general population. *Music Perception, 5,* 79-92.

Deutsch, D., North, T., & Ray, L. (1990). The tritone paradox: Correlate with the listener's vocal range for speech. *Music Perception, 7,* 371-384.

Hanley, T. (1951). An analysis of vocal frequency and duration characteristics of selected samples of speech from three American dialect regions. *Speech Monographs, 18,* 78-93.

Hanley, T. D., & Snidecor, J. C. (1967). Some acoustic similarities among languages. *Phonetica, 17,* 141-148.

Hanley, T. D., Snidecor, J. C., & Ringel, R. L. (1966). Some acoustic differences among languages. *Phonetica, 14,* 97-107.

't Hart, J., Collier, R., & Cohen, A. (1990). *A perceptual study of intonation: An experimental-phonetic approach to speech melody.* Cambridge, UK: Cambridge University Press.

Hermes, D. J. (1988). Measurement of pitch by subharmonic summation. *Journal of the Acoustical Society of America, 83,* 257-264.

Hollien, H., & Jackson, B. (1973). Normative data on the speaking fundamental frequency characteristics of young adult males. *Journal of Phonetics, 1,* 117-120.

Horii, Y. (1975). Some statistical characteristics of voice fundamental frequency. *Journal of Speech and Hearing Research, 18,* 192-201.

Houtsma, A. J. M., & Fleuren, J. F. M. (1991). Analytic and synthetic pitch of two-tone complexes. *Journal of the Acoustical Society of America, 90,* 1674-1676.

Houtsma, A. J. M., Rossing, T. D., & Wagenaars, W. M. (1987). *Auditory demonstrations CD.* Acoustical Society of America and Institute for Perceptual Research (IPO), Eindhoven, The Netherlands.

Hudson, A. I., & Holbrook, A. (1981). A study of the reading fundamental frequency of young black adults. *Journal of Speech and Hearing Research, 24,* 197-201.

Hudson, A. I., & Holbrook, A. (1982). Fundamental frequency characteristics of young black adults: Spontaneous speaking and oral reading. *Journal of Speech and Hearing Research, 25,* 25-28.

Jassem, W. (1971). Pitch and compass of the speaking voice. *Journal of the International Phonetic Association, 1,* 59-68.

Jassem, W., & Kudela-Dobrogowska, K. (1980). Speaker-independent intonation curves. In L. R. Waugh & C. H. van Schooneveld (Eds.), *The melody of language* (pp. 135-148). Baltimore, MD: University Park Press.

Liberman, M., & Pierrehumbert, J. (1984). Intonational invariance under changes in pitch range and length. In M. Aronoff & R. Oehrle (Eds.), *Language sound and structure* (pp. 157-233). Cambridge, MA: MIT Press.

Maeda, S. (1976). *A characterization of American English intonation.* Doctoral dissertation, MIT, Cambridge, MA.

Moore, B. C. J., Glasberg, B. R., & Peters, R. W. (1985). Relative dominance of individual partials in determining the pitch of complex tones. *Journal of the Acoustical Society of America, 77,* 1853-1860.

BEST COPY AVAILABLE

Mysak, E. Pitch and duration characteristics of older males. *Journal of Speech and Hearing Research*, 1959, 2, 46-54.

Ragozzine, F., & Deutsch, D. (1993). A regional difference within the United States in perception of the tritone paradox. *Journal of the Acoustical Society of America, 94*, 1860. (Abstract)

Shepard, R. N. (1964). Circularity in judgments of relative pitch. *Journal of the Acoustical Society of America, 36*, 2346-2353.

Snidecor, J. (1943). A comparative study of the pitch and duration characteristics of impromptu speaking and oral reading. *Speech Monographs, 10*, 50-56.

Terhardt, E. (1991). Music perception and sensory information acquisition: Relationships and low-level analogies. *Music Perception, 8*, 217-240.

Terhardt, E., Stoll, G., Schermbach, R., & Parncutt, R. (1986). Tonhöhenmehrdeutigkeit, Tonverwandtschaft und Identifikation von Sukzessivintervallen. *Acustica, 61*, 57-66.

Terhardt, E., Stoll, G., & Seewann, M. (1982a). Algorithm for extraction of pitch and pitch salience from complex signals. *Journal of the Acoustical Society of America, 71*, 679-688.

Terhardt, E., Stoll, G., & Seewann, M. (1982b). Pitch of complex signals according to virtual-pitch theory: Tests, examples, and predictions. *Journal of the Acoustical Society of America, 71*, 671-678.

Terken, J. (1993). Baselines revisited: Reply to Ladd. *Language and Speech, 36*, 453-459.

Willems, N., Collier, R., & 't Hart, J. (1988). A synthesis scheme for British English intonation. *Journal of the Acoustical Society of America, 84*, 1250-1261.

Yamazawa, H., & Hollien, H. (1992). Speaking fundamental frequency patterns of Japanese women. *Phonetica, 49*, 128-140.

FOOTNOTES

[*] *Music Perception*, 12, 227-255 (1994).

[1] Shepard did not consider the possibility that this lack of momentary ambiguity may be due to strong influences of preceding context. Informal observations by the author suggest that the interval C-F# is perceived as rising in the context of the interval sequence C-C#, C-D, C-D#, ..., but as falling in the context of C-B, C-A#, C-A, ... In fact, the author perceives *all* intervals up to C-B as rising in the first sequence and all intervals down to C-C# as falling in the second sequence, which suggests that sequential context effects can completely override the pitch proximity principle discussed by Shepard. (See Appendix B-1 for a discussion of sequential effects in the tritone task.)

[2] Deutsch made several changes with respect to Shepard's original stimuli: She used 6 rather than 10 octave-spaced partials, a somewhat differently shaped spectral envelope, longer tone durations (500 ms rather than 120 ms), and no silent intervals between successive tones in a pair. She did not explain the reasons for these changes.

[3] The comparison actually involved pairs of pitch classes: those straddling the limits of the speech octave band and those on top of the pitch circle. The tritone perception results were analyzed in such a way that two adjacent pitch classes emerged as "highest" (e.g., D# and E in Figure 1).

[4] Several more recent publications that summarize the same findings (Deutsch, 1992a, b, c) are equally nonspecific about the presumed link between "language" and tritone perception.

[5] It is difficult to find a direct comparison of British and American English in terms of F0 measurements. The Dutch researchers who have been most active in this area did not include American English in their studies.

[6] Actually, the average difference is less than one octave (see, e.g., Hudson & Holbrook, 1981).

[7] Two of the 12 tones had only 5 partials, as the first or sixth partial had zero amplitude.

[8] The envelopes were generated by substituting Deutsch's formula for the original Shepard (1964) formula in the program code. They look different from those in Deutsch's figures because they are drawn on a linear scale whereas Deutsch plots them on a dB scale. Some ripples in the curves are due to the fact that they were drawn by connecting 72 data points (6 partials x 12 stimuli), not by plotting the smooth mathematical function specifying the relative amplitudes.

[9] Except for their longer duration, these tones were identical with those recorded on a well-known CD of auditory demonstrations (Houtsma, Rossing, & Wagenaars, 1987).

[10] This was accomplished by first constructing a 12 x 12 matrix of the appropriate permutations of the numbers 1-12, and then assigning the 12 tone pairs at random to the 12 numbers.

[11] The tones ascended from D# to D in the Deutsch-D# and Shepard tone tests, and from A to G# in the Deutsch-A test. This rather unnecessary deviation from Deutsch's procedures will be discussed further below.

[12] This subject (MR-f, a violinist) was the only subject in this study known to possess absolute pitch.

[13] In one exceptional case (Dutch subject LB-m) a reversal of the pitch class effect in the middle of the Shepard tone test was noted. These data were scored according to the first half of the test, which seemed in better agreement with the results of the other subjects. Another Dutch subject (AB-f) gave only "up" responses in the first test (Deutsch-D#). She was encouraged to vary her responses, and when she repeated the Deutsch-D# test at the end of the session, she gave a clear pitch class effect. (See also Appendix B-3.) A few subjects yielded data that suggested a single highest pitch class rather than two.

[14] These instances were: Dutch subject JM-f in the Shepard tone test; British subject JS2-m in the Deutsch-D# test; British subject JC-f and American subject DK-m in the Deutsch-A test.

[15] The Californians were subjects DK-m, GS-m, SC-f, TM-f, MS-f, and HS-f. Other subjects had grown up mainly in the Midwest (JJ-m, JC-m, RM-m, MC-f), the East (ES-m, SL-m, AF-f, PK-f, MR-f), and the South (DW-m, CG-f); some of them had moved around as children. No information on subjects' parents was obtained. The author could not detect any relationship between subjects' regional origin and tritone perception results.

[16] These results were obtained by having listeners match pure tones to the perceived pitch of single Shepard tones. It would be worthwhile repeating this experiment, asking listeners to match pairs of pure tones to the perceived pitch change in tritone pairs. The author informally tried to match the tritone pitches by imitating them on a digital piano (i.e., with complex harmonic tones) while listening to portions of each test: Consistent with Terhardt's theory, he mapped the Deutsch-D# stimuli into the range $G\#_3$-A_4 (i.e., 208-440 Hz),and the Deutsch-A stimuli into the only slightly lower range $F\#_3$-G_4 (i.e., 185-392 Hz); Shepard tones, however, seemed to have much lower pitches, in the range A_1-B_2 (i.e., 55-123 Hz).

[17] A computational application of Terhardt's algorithm to the parameters of the present stimuli (Richard Parncutt, personal communication) has suggested that the effect of spectral envelope should be slight, in accord with Deutsch (1987). The large differences found in the present study await an explanation.

[18] As a regular subject in the Dutch group (77 dB presentation level), AH-m had A-A# as his highest pitch classes. Shifts of 1-2 st are hardly significant when the pitch class effect is relatively weak. American subject SC-f exactly replicated her earlier test results (cf. Figure 4).

APPENDIX A: READING MATERIALS

English sentences	*Dutch translations*
(1) I really felt sick yesterday, so I left work early and went home to lie down.	(1) Ik voelde mij zo beroerd gisteren, daarom ben ik vroeg naar huis gegaan en op bed gaan liggen.
(2) The three Baltic states gained their independence before the Soviet Union fell apart completely.	(2) De drie Baltische Staten verkregen hun onafhankelijkheid nog voor de Sovjet Unie volledig uit elkaar viel.
(3) Foreigners visiting the Netherlands often don't bother to learn the Dutch language.	(3) Buitenlanders die Nederland bezoeken, nemen vaak niet de moeite de Nederlandse Taal te leren.
(4) The concert last week was so boring that I dozed off during the performance.	(4) Het concert van vorige week was zo slaapverwekkend dat ik wegdoezelde tijdens de voorstelling.
(5) After the prince married the princess, they lived happily ever after.	(5) Nadat de Prins de Prinses had getrouwd, leefden zij nog lang en gelukkig.
(6) Among the endangered species in this world are elephants and tigers.	(6) Tot de bedreigde diersoorten in de wereld behoren de olifanten en tijgers.
(7) I attempted to call my friend three times, but the line was always busy and so I gave up.	(7) Ik heb drie keer geprobeerd mijn vriend te bellen, maar hij was steeds in gesprek en dus heb ik het opgegeven.
(8) Since I got myself a new bicycle it is much more fun riding to work in the morning.	(8) Sinds ik een nieuwe fiets heb, is het veel leuker om 's ochtends naar mijn werk te gaan.
(9) My mother was pleased to find that we had a large supply of paper towels in our cabinet.	(9) Mijn moeder was blij te zien dat wij een grote stapel papieren handdoeken in het keukenkastje bewaarden.
(10) Now that spring has arrived, the birds are singing and the trees are growing leaves again.	(10) Nu de lente eindelijk in het land is, zingen de vogels en lopen de bomen weer uit.

APPENDIX B: ADDITIONAL OBSERVATIONS

1. Sequential context effects

Imagine a subject who shows a perfect pitch class effect: 6 adjacent tritone pairs always receive "down" responses, the other 6 (their inverses) always receive "up" responses. This pattern of results (which was closely approximated by some subjects in some tests) implies sequential dependencies of the following sort: When a tritone pair is preceded by another pair whose pitches are 1 st higher or lower, it will receive the same response as the previous pair in 20 out of 24 instances. This is so because there are only 4 possible sequences of such pairs that cross the two sharp boundaries between "down" and "up" responses; all other sequences are within response categories. By the same reasoning, tritone pairs separated by 2, 3, 4, and 5 st will receive the same response in 16, 12, 8, and 4 instances, respectively, out of 24. Finally, pairs separated by 6 st (i.e., each pair followed by its inverse), of which there are only 12 instances, will never receive the same response. These sequential effects may thus be a consequence of the pitch class effect. If so, they should be absent when the pitch class effect is weak or absent.

Nearly all subjects tested showed strong pitch class effects on at least some tests, and those instances where the pitch class effect was weak may have been due to difficulties with the stimuli or the task. There was one listener, however, who consistently refused to show a pitch class effect, at least with the Deutsch tones, while being highly confident of his responses—the author (BHR). He served in extensive pilot runs with various stimuli and did show a weak pitch class effect initially during these runs. However, with the final sets of Deutsch tones, listening under the same conditions as the experimental subjects, he repeatedly gave flat response functions.

Figure B1 shows his results from one listening session. The top panel shows that there was no clear pitch class effect in either test. The bottom panel, however, reveals strong sequential effects—even stronger than would be implied by a perfect pitch class effect (diagonal line): BHR usually gave the same response to the current tritone pair as to the preceding pair when the pitch separation was 1-3 st, and different responses when the separation was 4-6 st. He hardly ever perceived any ambiguity in the direction of pitch change; yet he frequently reversed his responses to the same stimulus pairs. These results are proof, then, that sequential context effects can occur quite independently of the pitch class effect, and they lead one to wonder whether the latter might somehow depend on the former. The exact relation between the two phenomena is not clear at present.

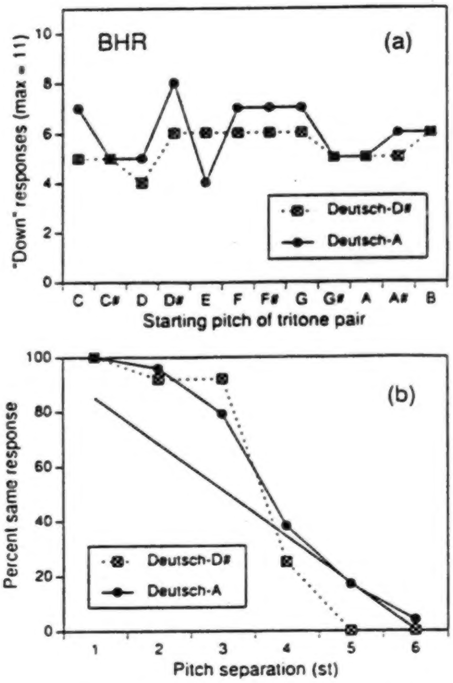


Figure B1. **Results of BHR for the two Deutsch tone tests, from a run conducted at IPO. (a) Response functions. (b) Percentage of identical responses to successive tritone pairs as a function of pitch separation. The diagonal line indicates the percentages implied by a perfect pitch class effect.**

2. Effects of presentation level

The lowest and highest partials of Deutsch tones are quite weak and probably contribute little to the virtual pitch percept. Still, it is conceivable that they have a disproportionate effect when they cross a threshold from inaudibility to audibility. The sudden appearance of low-frequency partials is actually quite striking as one listens to a regularly ascending sequence of Shepard or Deutsch tones, and this could be linked to the pitch class effect. If so, the pitch class effect should change with presentation level, since, with a fixed spectral envelope, presentation level must affect the pitch at which low-amplitude partials become audible.

Adrian Houtsma kindly provided the following data, which he collected near the end of the author's stay at IPO, using himself (AH-m) and an American colleague (SC-f) as subjects. (Both were also subjects in the present study; cf. Figure 4.) The test contained 10 repetitions of the 12

Deutsch-D# tritone pairs at each of 5 levels: 60, 65, 70, 75, and 80 dB SPL. The sequence was entirely random, and the subjects listened binaurally with Etymotic ER-2 insert earphones.

Figure B2 shows the results. Subject SC-f, on top, showed a very pronounced pitch class effect, with A#-B being the highest pitch classes; subject AH-m, below, had a less pronounced effect and noisier data, suggesting B-C as the highest pitch classes.[18] There is a suggestion in the data that the pitch class effect becomes weaker as presentation level decreases. Neither subject, however, showed any indication of a systematic shift in the response functions with presentation level. This suggests that low-amplitude partials do not play an important role in the pitch class effect, and that presentation level is not a critical variable in experiments on the tritone paradox.

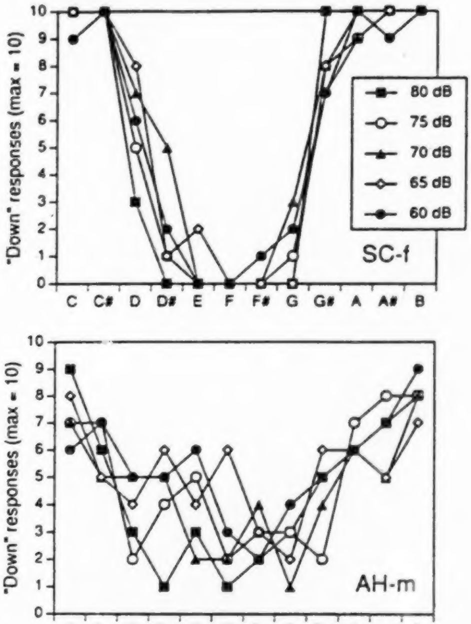


Figure B2. **Results of two subjects in a test using 5 different presentation levels (Deutsch-D# stimuli). Data courtesy of Adrian Houtsma.**

3. Listening strategies

Most subjects found the tritone tests easy and straightforward. A few, however, complained about ambiguity and claimed to hear occasionally simultaneous changes in opposite directions in different frequency regions. Presumably, these subjects were "analytic listeners" (cf. Houtsma and Fleuren, 1991), whereas most others were "synthetic listeners" who perceived only a single dominant pitch and little ambiguity. Each of these analytic listeners nevertheless produced consistent pitch class effects, apparently by adopting a consistent strategy. Their comments suggested, however, that their results might have been different if they had adopted a different strategy. (See also Footnote 13.) This possibility was checked out by recalling three Dutch subjects (DH-m, WR-m, AB-f) who had complained about the ambiguity of the tritone pairs. In this follow-up test, they were presented twice with the Deutsch-D# test and were asked to listen to the low frequencies the first time and to the high frequencies the second time.

The results of two subjects, DH-m and WR-m, are shown in Figure B3. To the author's surprise, their pitch class effects were not affected by the change in listening strategy. (The very small differences visible in the figure would require more extensive data to be considered significant.)

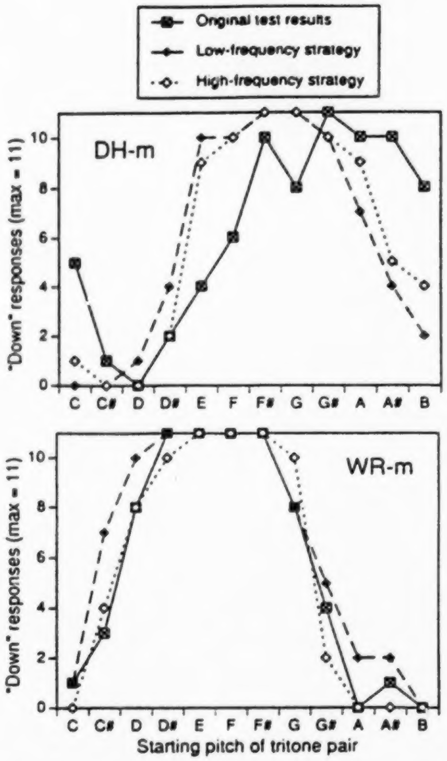


Figure B3. **Results of two subjects employing different listening strategies in the Deutsch-D# test. Their original test results are also shown.**

Subject WR-m closely replicated his results of the original Deutsch-D# test run, whereas subject DH-m appeared to have shifted his pitch class effect down by 2 st. The third subject, AB-f, showed a pitch class effect with the low-frequency strategy which, too, was 2 st below her original results, but she gave only "up" responses when listening to high frequencies. (See also Footnote 13.) Subsequent questioning revealed that she had been listening to very high frequencies, whereas the other two subjects listened somewhere in the middle region, near the strongest partials. The low-frequency listening strategy apparently focused on the lowest virtual pitches or audible partials of the tone complexes; DH-m called them "the bass notes".

Why, then, did these subjects claim to hear opposite pitch changes in different frequency regions? Presumably, these impressions derived from the ambiguous tritone pairs in the transition zone between "up" and "down" responses. If the responses to these pairs were reversed, this would not change the pitch class effect. The detailed response protocols of subject WR-m were examined to determine the number of response reversals. Between his low-frequency (LF) and high-frequency (HF) strategy runs, there were 22 reversals (out of 132 scored responses). Between the LF run and the original test, there were also 22 reversals, but between the HF run and the original test there were only 12. This suggests that WR-m employed the HF strategy in the original run, which seems plausible. The finding that there were almost twice as many reversals with the LF strategy may then be taken as evidence in support of the hypothesis that responses to ambiguous tritone pairs were affected by listening strategies. Subject DH-m, however, showed only 14 response reversals between his LF and HF runs, and because of his shift in the pitch class effect, a count of reversals with respect to the original test results would not have provided a proper comparison. Therefore, these results remain suggestive at best.

Haskins Laboratories Status Report on Speech Research
1993, SR-115/116, 145-147

A Review of Treiman, R. (1994). *Beginning to Spell**

Donald Shankweiler†

Less is known about how people learn to write than about how they learn to read. Spelling is often downplayed in discussions of literacy. Reasons are not hard to find. In American society at large, even among those who teach, the spelling system of English is widely disparaged as so hopelessly irregular as to scarcely be worth the effort required to master it. Rebecca Treiman reminds us, however, that spelling is an important part of literacy. Just as fluent reading of words is critical to skillful reading comprehension, so learning to spell words is important for attaining fluency in writing. Treiman's book does much to fill the large gaps in our knowledge of how young children come to grips with English spelling. It is an important book for all who study literacy—psychologists, linguists. educators. With the call for more emphasis on writing during the early grades, this book contains much valuable information that could prepare teachers to respond intelligently to children's mistakes. It is also an important book for psycholinguistic researchers on child language for the insights and data it contains on how children represent the sound structure of their language at the onset of literacy.

Ironically, young children may be more willing to treat spelling as a system than many of their elders. We learn from this book that children in the first grade of an American. mid-western elementary school. long before they have been taught the intricacies of English spelling, and armed with only a rudimentary knowledge of the alphabet, display remarkable ingenuity and resourcefulness in inventing spellings for many of the words they used in writing stories. These spellings often do not conform to the standard system, to be sure. but when analyzed with the insight Treiman brings to the task, they become intelligible and are shown often to make good sense linguistically. A child who writes SGIE for *sky* is displaying acute phonological judgment by using the letter *g*, which usually represents an unaspirated velar stop, in preference to *k*, which commonly represents an aspirated velar. A child who writes AVR for *ever* is using the letter a to represent a vowel that in fact is sometimes represented by that letter (as in *bare).* Moreover, this child has arguably not omitted the vowel in the second syllable, but has represented it by making the letter r do double duty for both vowel and consonant. The letter's name would surely encourage just that error. The many interpreted examples in this book show that, to penetrate beneath the surface of children's spellings, one must look beyond the mere fact that a spelling is right or wrong. Instead. the questions become: Which of the word's phonemes are represented. and how are they represented: which are omitted, and why?

Take an ordinary classroom of first graders in an ordinary public school. Give them a teacher who requires them to write regularly but who does not (at this stage) correct their spellings. If they are like the children Treiman studied, they will work intelligently to apply the alphabetic principle as they understand it. What the young writers most often capture in their spellings, Treiman argues. is their conceptions of the phonological structure of words. Frequently, the children generate much the same kinds of representations as the conventional orthography. Where the spellings are aberrant. these reflect the children's lack of experience, not totally wrong guesses about how the alphabet represents the language.

The idea that young children's spellings are chiefly an attempt to represent the phonological forms of words is the central theme of this book. Over and over, Treiman's findings demonstrate that, once past the earliest pre-alphabetic stage, children's spellings honor the alphabetic principle: they differ from conventional spellings because the structural representations of words in the children's mental lexicons are incomplete. and because children at the beginning stages of reading and writing have very limited knowledge of the orthography's resources. Experience in reading and writing will continue to shape their phonological representations.

The findings Treiman reports do much to dispel widespread misconceptions about the nature of spelling skill. The conventional wisdom holds that spelling is chiefly a visual memory task, learned passively and largely by rote. It is difficult to square these assumptions with many of Treiman's findings. "At least three processes seem to be involved in spelling a word: analyzing the spoken word into smaller units. remembering the identity and order of the units. and assigning a grapheme to each unit" (p.280). The first process requires phonological awareness. Ultimately, the writer must appreciate that words come apart into phoneme segments. The second involves memory. but not primarily visual memory. Children must store the units in phonologic short-term memory while carrying out the other processes. They must remember both the identity of the units and their order. Finally. in order to transcribe each unit as a letter or group of letters, they must use their stored knowledge of the correspondences between phonemes and spellings. It is apparent from this conception of spelling that far more is involved than recall of visual shapes. Misspellings may reflect several linguistically-driven processes.

The insight that children's early efforts to write often reveal penetrating attempts to represent the internal structure of words will be known to readers who are familiar with the literature on children's invented spellings stemming from the researches of Carol Chomsky and Charles Read. The studies reported in this book build on their discoveries, particularly Read's work. as Treiman notes repeatedly. Moreover. the research builds on the concept of phonological awareness, long recognized as critical for mastery of the alphabetic principle. This book presents some of the strongest evidence to date that skill in spelling. no less than in reading, rests on phoneme awareness. Even after a child apprehends that words come apart into segments. however. it may be difficult 10 apply this insight in all syllable contexts. Thus. the strong tendency of beginners to simplify consonant clusters, for example. spelling *trap* as TAP, may reflect the difficulty in apprehending the sound represented by *r* as an separable element of the word when it is part of a cluster.

The value of the book is enhanced by inclusion of a generous early portion devoted to theoretical background. With an eye to the needs of readers who are not trained in linguistic phonology,. Treiman presents the linguistic concepts required to grasp and interpret the research findings. Although the book is apparently designed to be self-teaching, this material, and, indeed, the findings themselves, will no doubt be appreciated most by those who have already assimilated some of the background. Unfortunately, not everyone recognizes the need for specialized knowledge of phonology and other branches of linguistics and psycholinguistics as preparation for studying how people acquire basic literacy skills. Where language is concerned, it seems as though nearly every schooled adult regards himself or herself as already an expert. Treiman effectively counters this attitude by demonstrating how indispensable are certain linguistic concepts and tools for understanding how children arrive at the spellings they produce.

To summarize the book's special strengths: First, its thoroughgoing linguistic approach, relating the problems of spelling to the phonological and morphological structures of the language and children's apprehension of those structures. It assesses how children spell in an actual creative writing situation: Words are generated in the context of stories the children produce. The analysis of the data takes account of correct spellings as well as errors, so that it is possible to discover which types of words are easiest for beginners to spell correctly. Treiman's database is more representative than Read's: her subjects were from middle-class backgrounds and attended a public school. They were not children with a precocious interest in writing, and they did not come chiefly from the most highly-educated families. Some of the limitations of a naturalistic study are overcome by the inclusion of full discussions of the author's earlier experimental work, which provided much of the empirical and theoretical framework for this book.

In a study of this kind, there are. unavoidably, some limitations. The data-collection method has drawbacks as well as virtues. as Treiman freely acknowledges. Perhaps the most serious drawback is that there is no control over the words attempted (and not attempted) by the children. Contributions to the corpus by different children were very unequal. On the whole, individual differences get short shrift in this book. We find little about possible effects of the child's reading strategy. So, for example, we do not learn whether "logographic" readers have a discernibly different approach to spelling than "analytic" readers. Treiman does consider the important question of how reading experience may influence spelling, but. as she notes. the materials she had available to her the children's collected writings) did not allow her to investigate this question, nor could she study the possible influences of spelling on reading. For some of the questions considered in this book. a cross-language perspective could be helpful. One cannot easily sort out the separate contributions that language and orthography make to spelling difficulties without recourse to comparative data. Though it is packed with new content throughout. the book could have benefited from shortening. A certain amount of redundancy is guaranteed by the organization: Each chapter is self-contained, and each includes a detailed summary that includes recaps of portions of the theoretical argument.

Overall, Treiman's book is major achievement in research on literacy acquisition. It yields a rich harvest of new findings. and confirmation of some important old ones. Indeed. this book has raised the study of spelling to a new level of conceptual and methodological sophistication. It will surely become the standard by which future work is judged for a long time to come.

FOOTNOTES

[*] New York: Oxford University Press, 1993. 365pp. $49.95. This review appears in *Language and Speech, 37(1), 77-79* (1994).

[†] Also University of Connecticut, Storrs.

Haskins Laboratories Status Report on Speech Research
1993, SR-115/116, 149-153

A Review of McNeill, D. (1992). *Hand and Mind: What Gestures Reveal About Thought* *

Michael Studdert-Kennedy

The argument of this original and difficult book is that "gestures are an integral part of language as much as are words, phrases and sentences—gestures and language are one system" (p.2). Gestures are instantaneous, imagistic, analog, holistic expressions of the same thought that speech renders in hierarchical, linear, digital, analytic form. David McNeill credits Adam Kendon (1972, 1980) with discovering the link between, and essential unity of, speech sounds and gestural movements; his own work elaborates this insight at the higher linguistic levels of semantics and pragmatics.

The topic of the book, then, is gestures that accompany speech, the left-hand end of what McNeill calls "*Kendon's continuum*: Gesticulation ⟶ Language-like Gestures ⟶ Pantomimes ⟶ Emblems ⟶ Sign Languages" (p.37). The continuum ranges from the informal, spontaneous, idiosyncratic movements of the hands and arms that often accompany speech, to the socially-regulated, standardized, linguistic forms of a sign language, with its arbitrary (non-iconic) lexicon.

Between these poles the obligatory presence of speech declines and the linguistic properties of gestures increase. "Language-like gestures" are grammatically integrated into an utterance, as when a speaker, asked about the weather on his vacation, replies: "Well, it was [oscillating hand gesture]," where the "so-so" gesture replaces an adjectival predicate. "Pantomime" conveys its full meaning in silence or, at most, with inarticulate onomatopoeia; also, in pantomime, sequences of gestures can form a unit, as they can in a sign language, but cannot in gesticulation. "Emblems" conform to standards of well-formedness, a language-like property that gesticulation and pantomime lack: in England, the palm-front V-sign is Churchill's "Victory!," the palm-back V-sign is a sexual insult. (For an amusing cross-class confusion in emblem dialects, see Morris, Collett, Marsh, & O'Shaughnessy, 1979, p.229, where Margaret Thatcher appears in an Associated Press Photo, making the palm-back V-sign at a moment of electoral triumph.)

The contrast between the two ends of Kendon's continuum, between spontaneous gesture and conventional sign, epitomizes McNeill's notion of the process by which an utterance evolves in a speaker's mind. Spontaneous gesture reveals the primitive stage of an utterance, global, unsegmented, non-hierarchical, from which its conventional representation in speech unfolds: hierarchical, segmented, linear. The inner symbols of the primitive stage are private, idiosyncratic, closed to social influence; the end stage is public, grammatical, socially regulated. McNeill supposes that the primitive stage of a sentence in a conventional sign language, such as American Sign Language (ASL), consists of global images no less than in a spoken language. But these images cannot escape into gesture, because the public end state has preempted the gestural channel.

The microevolution of an utterance may, in its turn, epitomize the macroevolution of linguistic system from primitive gesture. To support this speculation, McNeill reports the results of an undergraduate thesis at the University of Chicago by Ralph Bloom (1979). Bloom videotaped adults, who had no knowledge of a sign language, while they recounted a traditional fairy story to an adult viewer without speaking. Striking changes took place over the course of a session. For example, within 15 minutes or so, one story teller had developed a system with many of the standard properties of a spoken or signed language: segmentation; compositionality (of both signs and propositions); a lexicon (including three types of pronoun, one of them an abstract spatial pronoun similar to those of ASL); paradigmatic opposition (as when "King" and "Queen" shared a hand circling the head for "crown," but were distinguished by iconic gestures for "has-muscles" vs. "has-breasts"); ergativity (for example, the incorporation of a noun into a verb, as when the movement of threading a needle is made with a hand shaped for holding a thread); sign or "word" order (usually SV or SVO); standards of well-formedness; and fluency. With regard to the last, not only did the storyteller begin to streamline signs, stripping them to their essential features, but the viewer became so comfortable that he began to formulate questions in the new "language."

Although some of these linguistic properties may have been modeled on English (e.g., "word" order), others evidently were not modeled on any language known to the storyteller (e.g., three types of pronoun, ergativity). Rather, they seem to have emerged automatically, shaped by pressures toward clear and expressive communication. An analogous process in ontogeny, the emergence of recursion in the signing of a deaf child of hearing parents who used no sign themselves, has been reported by Goldin-Meadow (1982).

Whatever the worth of these parallels, or of the microgenesis of an utterance itself, as models of the phylogenetic evolution of language, they emphasize two aspects of Kendon's continuum essential to McNeill's argument. First is the contrast between spontaneous gesture (gesticulation) and formal language, whether spoken or signed. Second is the central theme of the book, namely, the common origin of the contrasting modes, gesture and language, in the thought that a speaker intends to express.

Let us turn now to the empirical work on which McNeill's monograph is based. Most of the gestural examples come from quasi-experimentally induced narrative discourse, although some are drawn from TV broadcasts, videotaped conversations, and naturally observed academic discourse. In the basic experimental situation, a speaker sees a film, animated cartoon, or comic book, and then recounts its story to a listener, a genuine listener who has not seen the "stimulus," and who will later have to retell the story to a third person. Neither speaker nor listener knows that gestures are the objects of study. The entire session is recorded on audio-video tape and the tape is subjected to minute analysis. All spoken utterances are transcribed, clause by clause, together with an indication of hesitations, and of the durations of pauses, filled and unfilled, in tenths of a second. All gestures are classified, coded and transcribed, with an indication of the words, parts of words, or pauses with which they were temporally aligned. (A 25-page Appendix includes instructions for coding and transcription, detailed enough for other researchers to replicate the procedures of McNeill and his associates.)

The four main types of gesture in McNeill's system are: iconics, metaphorics, beats and deictics. All are symbolic, in that hand and arm stand for something other than themselves, and all are closely related to the semantic and/or pragmatic aspects of the speech they accompany. Iconics are "gestures of the concrete," exhibiting more or less transparent images of their referents. The images may be redundant, that is, coexpressive, with speech—a circling hand with downward pointing index finger for a cake on a table, a rising hand for someone climbing—, or complementary, capturing an aspect that the speech misses, as when a narrator describes an old woman chasing a cat out of her house and indicates her weapon, an umbrella, not in words, but with threatening shakes of the forearm.

Complementarity is one of several gestural properties demonstrating that gesture is neither a different version of the same covert verbal plan as speech, nor a translation from speech into another modality, nor, finally, an independent visual display, a photograph as it were, of the scene that speech puts into words. Rather, "...gestures and speech are operations that have been connected *within*" (p.33, italics in the original), each arising from the same emergent thought, separate, yet integral, and each essential to full expression of a speaker's meaning.

Another index of this relation is the use of iconic gesture to highlight what a narrator finds salient in a situation. For example, a narrator describes two attempts by a cartoon cat to climb a drainpipe, first up the outside, then, having failed, up the inside: for the first attempt the narrator's hand rises with the palm flat, for the second, with the palm in a hollow basket shape, depicting the interiority of the path. Here, the gesture reveals the speaker's "psychological predicate" (Vygotsky, 1962) at the moment of speaking, that is, "the novel, discontinuous, unpredictable component" (p.127), that sets her current thought off from what went before. Evidently, the aspect of the second climb that the speaker particularly wished to capture was that it was inside rather than outside the pipe. If, as Vygotsky argued, thought is a process of forming contrasts with respect to preceding context, and if gestures express these contrasts (or "psychological predicates"), then we may sometimes apprehend the full meaning of an utterance more clearly, or even only, by attending to gestures as well as speech.

Not surprisingly, iconics tend to predominate in narrative, but "gestures of the abstract" (metaphorics, beats, deictics) also often occur and predominate in other genres, especially conversations and lectures. Metaphorics are no less pictorial than iconics, but the image they present is of an abstraction. For example, a type of metaphoric, called a "conduit metaphor," represents language, meaning, knowledge, art or other abstract notions as a substance, packed into a container that can be passed from one person to another. (Such metaphors are common in speech: "empty words," "deep book," "an amusing article, but not much in it," "a difficult idea to get across," and so on.) Thus, a speaker introduces a narrative with the words: "It was a Sylvester and Tweety cartoon"; as he speaks, he raises his hands as though holding a box, and then moves them apart, as though breaking it open. In another example, a speaker says: "I have a question," holding out her hand in a cup shape, as though to receive an answer.

The "cup-of-meaning" handshape seems to be a common symbol, adaptable to diverse circumstances. A speaker, referring to an event that might have happened but did not, says: "...even though one might have supposed..."; with the first four words his hands move out to the side in a cup shape, symbolizing potentiality (waiting for something to fall into them), on the fifth word ("have"), the hands snap shut onto emptiness (nothing fell after all). Or again, a speaker at an academic conference, emphasizing the importance of organization in a certain domain, moves his hand forward in a cup shape, as though carrying the domain itself, then with the word "organization," abruptly extends and spreads his fingers to form a rigid supporting armature.

Many other nicely analyzed examples—metaphors for states of mind, for dynamic processes of change, for mathematical concepts, and so on—illustrate the ease with which abstract ideas take on gestural form. Evidently, a concrete image aids the expression, and perhaps communication, of abstract thought.

The two remaining major types of gesture (beats, deictics) are not pictorial. Beats are simple two-phase movements (in/out, up/down) in which the hand moves rhythmically in time with the speech, typically taking the same form regardless of context. Beats fulfil a pragmatic function within a discourse, indicating that a word or phrase is important, not for its own semantic content, but for its contribution to the development of the narrative, or argument. Beats typically accompany a change of scene, the introduction of a new character or of a new theme.

Finally, deictics (pointing gestures) may be either concrete or abstract. The abstract function is particularly striking in light of the extensive formal use of deixis in ASL (Klima & Bellugi, 1979). Here, of course, we are concerned with the spontaneous, often idiosyncratic, informal use from which ASL forms presumably arose. A simple example comes from a conversation in which a speaker asks: "Where did you come from before?," pointing to the space between himself and his hearer. The indicated space is clearly not the actual space between the two interlocutors, but an abstract concept of the place that the hearer came from. In a more elaborate example, a narrator adopts the space in front of him as a metaphor for the plot of a story, assigning different loci to different characters and different modes of action. Appropriate pointings to one locus or another then gesturally lead the listener through the story.

Before we leave the empirical work, we must briefly review one more class of evidence supporting the hypothesis that speech and gesture arise as separate, but integral, expressions of the same thought—namely, the relative timing of the two forms: gesture and speech have a constant temporal relation. A prototypical gesture has three phases: preparation, the "stroke," or main part, of the gesture, and retraction. Only the stroke is mandatory, and only the stroke is matched, or "synchronized," with the speech at three levels: phonological, semantic and pragmatic. The stroke precedes or ends at, but never follows, "...the phonological peak syllable of speech" (p.26). (Presumably this refers to the syllable that carries primary sentence stress in the judgment of a listener, although McNeill leaves "phonological peak" undefined.) The two channels also simultaneously express the same meaning ("semantic synchrony") and/or, where relevant, fulfil the same pragmatic function ("pragmatic synchrony").

Of particular interest here is the optional preparatory phase. For example, a narrator describes a comic book character in action: "...he grabs a big [oak tree and he *bends it way back*]'. (The square brackets enclose the words accompanying the gesture.) The preparatory phase begins with "oak": the speaker's hand rises up and forward at eye level, taking on a grip shape. Over the stroke phase (italicized words), the hand appears to pull something back and down toward the shoulder, ending at the phonological peak with the word "back." Now, since the preparatory phase has no function other than to prepare for the stroke, we can infer that the image of bending the tree back was already taking shape when the speaker was saying "oak tree." Evidently, then, the speaker's thought, the "minimal idea unit," or "starting node," from which both utterance and gesture grow had already taken global form during the preparatory phase of the gesture, before either the gestural stroke or the linguistic structure that would jointly express the thought had begun to emerge.

In such analyses as this (of which the book contains many dozens) we see with what subtlety McNeill builds, from empirical data, a theory of the relation between thought and its expression. His theory carries forward the work of Vygotsky (1962, 1986), deepening its biological roots and deliberately challenging (although quite without polemic) the programmatic (non-empirical), mechanical models favored by information-processing psycholinguists (e.g., Levelt, 1989).

For McNeill the starting node of an utterance (what we called above its "minimal idea unit" or "psychological predicate") is its "growth point," a metaphor from embryology with dynamic implications that its alternatives lack. The growth point is a small deviation, a minor salience, among the disordered fragments of images and linguistic categories, the residue of immediately preceding thoughts, from which utterance and gesture assemble themselves, thus assuring some degree of sequential coherence. The deviation does not arise tautologically from thought itself, but from extra-cognitive "...disruptive forces—motivation, emotion, and a future orientation" (p.239) within the speaker. Here McNeill quotes Vygotsky: "Thought is not begotten by thought; it is engendered by motivation, i.e., by our desires and needs, our interests and emotions. Behind every thought there is an affective-volitional tendency, which holds the answer to the last "why" in the analysis of thinking" (Vygotsky, 1986, p.252).

The assembly of utterance and gesture, set in motion by the "affective-volitional tendency," is a self-organizing process through which global order arises from local interactions among the disordered fragments, a theoretical process "...inspired by self-organizational models in developmental neurobiology (von der Malsburg and Singer, 1988)." Because a growth point consists of both images (which will surface as gestures) and linguistic categories (which will surface as words or phrases), the process of self-organization is a dialectic between gesture and language. The relation between the two modes changes during the few seconds of utterance-gesture formation (an interval that McNeill calls "deep time"): Thought, in its primitive stage more imagistic than analytic, emerges from deep time into real time, and so into existence, as a synthesis of the two modes. Essential to the synthesis is the underlying rhythmic pulse, the point of temporal convergence at which speech and gesture are integrated. With each pulse we, speaker and listener, gain momentary access to the "endless braid" (p.237) of thought that constitutes a life.

Perhaps I have now said enough to give the reader a sense of what this book is about, and of where it stands in the field of contemporary cognitive psychology. I have touched on its main themes, but have omitted many important topics, including those covered by two fascinating final chapters on the development of gesture in children and on the cerebral control of gesture, as evidenced by studies of aphasic and split-brain patients.

I have also omitted discussion of two chapters, reporting various experimental tests of the theory. One of these chapters describes several ingenious experiments, with delayed auditory feedback and other techniques, that support aspects of the self-organization model. The other chapter addresses questions that I found myself asking many times as I labored through the early chapters: Do gestures have any communicative function? Do listeners really use them to pick up information?

A partial answer comes from an experiment in which the "stimulus" for a speaker's narration of a cartoon film was not the film itself, but an experimenter's narration of the film, in which certain gestures had been deliberately mismatched to their accompanying speech. For example, the experimenter first established his left side as the locus of Sylvester, the wicked cat, and his right side as the locus of Tweety Bird, the good canary. Later, describing how Sylvester made a grab for Tweety the experimenter used anaphoric pronouns of which the verbal context made the reference unambiguous ("he lunges for him," where he = Sylvester), but accompanied these words with a gesture of his right hand (= Tweety) lunging to the right. The mismatch was subsequently reflected in the subject's verbal account: she initially reported in words what she had seen in gesture, while making a gesture matched to what she had heard in words; then she corrected herself, repairing her speech and repeating her original, correct gesture. Here, then, not only did gesture affect the subject's representation of the event, but it affected her initial recounting of it in words, while the words she heard determined the form of her gesture. The outcome not only demonstrates that gesture can convey meaning, but illustrates the unconscious integration of speech and gesture into a single "cross-modal" representation.

Nonetheless, my doubts concerning communicative function remain. Evidently gestures *can* convey information, and perhaps often do. Yet no less often gestures express essentially the same information as speech, and are therefore redundant. Moreover, listeners over radio or telephone do not seem to be especially handicapped. On the other hand, speakers over the telephone often gesture very much as they do in face-to-face conversation. And this suggests that gestures, even if not "cognitive necessities" (p.259), may nonetheless facilitate verbal expression, as indeed McNeill argues.

I came to see, however, as I worked through the book, that the communicative function of gestures is largely peripheral to McNeill's enterprise. What is central, once again, is the hypothesis that gesture and language are part of a single system, a hypothesis for which the book marshals a mass of compelling evidence. For it is from this insight that McNeill has developed a unique, empirically based, and biologically driven theory of the relation between thinking and speech. His approach is a refreshing change from the machine models that dominate cognitive psychology in the current social climate of the military-industrial complex. Moreover, by viewing the "primitive stage" of thought as imagistic, and gesture as its natural mode of expression, McNeill throws light on one of the most remarkable discoveries of modern linguistics: the equivalence, in function and abstract form, of spoken and signed language.

In conclusion, let me strongly recommend a book that I initially found tedious and difficult, but ended by admiring and enjoying. The difficulty and tedium stemmed partly from the novel descriptive categories, couched in an utterance-gesture notational system that I did not know, partly from the meticulous density of the argument, and partly from my uncertainty as to where all the dry *explication de geste* was leading. My admiration and enjoyment grew as I gradually apprehended the originality of McNeill's theory and the scope of its implications for the evolution of thought and language.

REFERENCES

Bloom, R. (1979). *Language creation in the manual modality: A preliminary investigation*. Bachelors thesis, Department of Behavioral Sciences, University of Chicago.

Goldin-Meadow, S. (1982). The resilience of recursion: A study of a communication system developed without a conventional language model. In E. Wanner & L. Gleitman (Eds.), *Language acquisition: The state of the art* (pp. 51-77). Cambridge: Cambridge University Press.

Kendon, A. (1972). Some relationships between body motion and speech. In A. Siegman & B. Pope (Eds.), *Studies in dyadic communication* (pp. 177-210). New York: Pergamon Press.

Kendon, A. (1980). Gesticulation and speech: Two aspects of the process of utterance. In M. R. Key (Ed.), *The relation between verbal and nonverbal communication* (pp. 207-227). The Hague: Mouton.

Klima, E., & Bellugi, U. (1979). *Signs of language*. Cambridge: Harvard University Press.

Levelt, W. J. M. (1989). *Speaking: from intention to articulation*. Cambridge, MA: M.I.T. Press/Bradford Books.

Morris D., Collett, P., Marsh, P., & O'Shaughnessy, M. (1979). *Gestures: Their origins and distribution*. New York: Stein and Day.

von der Malsburg, C., & Singer, W. (1988). Principles of cortical network organization. In P. Rakic & W. Singer (Eds.), *The neurobiology of neocortex* (pp. 69-99). Chichester, England: John Wiley.

Vygotsky, L. S. (1962). *Thought and language* (E. Hanfmann and G. Vakar, Trans.) Cambridge, MA: M.I.T. Press.

Vygotsky, L. S. (1986). *Thought and language* (A. Kozulin, Ed.). Cambridge, MA: M.I.T. Press.

FOOTNOTE

*Chicago University Press, 1992. xi, 416 pp. $34.95. This review will appear in *Language and Speech*.

Haskins Laboratories Status Report on Speech Research
1993, SR-115/116, 155-156

A Review of Lieberman, P. (1991). *Uniquely Human**

Michael Studdert-Kennedy

The subtitle of this book is: The *evolution of speech, thought and selfless behavior*. Its thesis is that "...the 'key' to the evolution of the modern human brain is rapid vocal communication" (p.9), afforded by "encoded" speech and syntax; "moral progress...follows from our cognitive ability which, in turn, derives from our linguistic ability" (p.10). The book consists of an introduction and six chapters, reviewing a wide range of data and theories concerning human brain structure, speech and syntax, aphasia and other language deficits, language acquisition, and the emergence of human culture.

The longest chapter treats topics that Philip Lieberman (hereinafter, L) knows best and for which he is best known, namely, speech physiology and the evolution of the human vocal tract. L's studies in this area are not without their critics, but anyone who does not know them would profit from reading the relevant sections of Chapter 2 (pp. 53-77). Unfortunately, I cannot say as much for the rest of the book, which has many defects of both substance and scholarship.

Substantively, L asserts rather than demonstrates the supposed evolutionary line from speech to thought to "selfless behavior", leaning on several tenuous assumptions; these include a misleading parallel between structure-dependent syntactic rules and context-dependent phonetic variation (e.g., p.83). More generally, despite its subtitle, the book is less concerned with evolution than with its presumed products: the human brain, vocal tract, and certain cognitive capacities. L has nothing to say about the properties of linguistic *behavior*, the "pacemaker" that must have driven the evolution of human morphology (Mayr, 1982, p.612), nothing to say therefore about the emergence of the discrete elements of sound and meaning, and their combinatorial structures, that afford language its unlimited semantic scope. These topics are no longer completely intractable (see, for example, Bellugi & Studdert-Kennedy, 1980; Lindblom, 1986; Pinker & Bloom, 1990), and should surely be considered in a book purporting to address the evolution of speech.

As for scholarship, L repeatedly fails to acknowledge the sources of ideas he wishes to promote, and to represent fairly those he does not. Consider, first, a notion at the center of L's argument, the so-called "encoding" and "decoding" of speech and syntax. L notes that we commonly produce and perceive speech at rates as high as 15-25 phonetic segments per second. "This fact leads to a seeming mystery...[because Miller (1956)]...showed that humans cannot identify non-speech sounds at rates that exceed seven to nine items per second.... How, then, can we possibly understand speech...?" (pp. 37, 38; cf. p.59). The answer, L tells us, is that we have specialized neural mechanisms for "encoding" and "decoding" speech and syntax. He draws a parallel between "encoded" speech which evades limits on the temporal resolving power of the ear, and "encoded" syntax which evades limits on short-term memory (p.82). Finally, he proposes that brain mechanisms underlying the "...complex muscular maneuvers of speech may have provided the preadaptive basis for rule-governed syntax" (p.83).

Several things are wrong here. First, Miller's (1956) paper deals with limits on the channel capacities of human perceptual systems and has nothing whatever to say about rate of processing. In fact, the speech rate puzzle was first remarked by Liberman et al. (1967), who also introduced the concept of "encoded" speech, contrasting a cipher, such as the alphabet, which substitutes a single symbol for each unit of the message, with a code in which message units are "restructured": Speech was said to be a code because its segments are merged into larger syllabic units by coarticulation, thus evading limits on the temporal resolving power of the ear. Later, Liberman (1970) drew a parallel between the interleaved patterns of speech and syntax, arguing for analogous specialized decoding devices for phonology and syntax. Finally, Lenneberg (1967, Chapter 3) elaborated at length the possible homologies between syntax and motor control. L's failure to credit these ideas to their sources indicates, at the least, extraordinarily careless scholarship (not to mention incompetent reviewing by his publisher, Harvard University Press).

Consider, next, L's treatment of Chomsky's universal grammar (UG), which he has quite evidently not taken the trouble to understand (see particularly pp. 127-134). L charges that UG is "biologically implausible", because it admits of no genetic variation. Yet in a work that L cites, Chomsky (1986) states: "...UG is a species characteristic, common to all humans. We...abstract from possible variations among humans in the language faculty.... Apart from pathology (potentially an important area of inquiry) such variation as there may be is marginal..." (p.18). Here and elsewhere (e.g., Chomsky, 1982, pp. 24,25), Chomsky acknowledges variation but, following standard biological practice, proposes a general, species-specific characteristic.

In another misguided passage, L assures us that specialized brain modules (currently favored by many neuropsychologists: e.g., Gazzaniga, 1989) are incompatible with the "'mosaic' principle" of evolution (p.6). L mistakenly identifies the principle with the determination of the several parts of an organ (the knee socket is his example) by independent genes. If he were correct, we would have no complex, polygenically determined organs at all. In fact, mosaic evolution refers to the "...highly unequal rates of evolution of different structures and organ systems..." (Mayr, 1982, p.613), and is fully compatible with brain modules.

I have had space to illustrate only a few of L's numerous errors. *Uniquely Human* is a curious compendium of fact and fiction, representation and misrepresentation, understanding and misunderstanding—in short, uniquely Lieberman.

REFERENCES

Bellugi, U., & Studdert-Kennedy, M. (Eds.) (1980). *Signed and spoken language: biological constraints on linguistic form.* Deerfield Beach, FL: Verlag Chemie.

Chomsky, N. (1982). *The generative enterprise.* Dordrecht, Holland: Foris Publications.

Chomsky, N. (1986). *Knowledge of language.* New York: Praeger.

Gazzaniga, M. (1989). Organization of the human brain. *Science, 245,* 947-952.

Lenneberg, E. H. (1967). *Biological foundations of language.* New York: John Wiley.

Liberman, A. M. (1970). The grammars of speech and language. *Cognitive Psychology, 1,* 301-323.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., and Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review, 74,* 431-461.

Lindblom, B. (1986). Phonetic universals in vowel systems. In J. J. Ohala, & J. J. Jaeger (Eds.), *Experimental phonology* (pp. 13-44). New York: Academic Press.

Mayr, E. (1982). *The growth of biological thought.* Cambridge, MA: Harvard University Press.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63,* 81-97.

Pinker, S., & Bloom, P. (1990). Natural language and natural selection. *Behavioral and Brain Sciences, 13,* 707-784.

FOOTNOTE

*Cambridge, MA: Harvard University Press. 1991. pp. 210. Paper. $12.95. This review will appear in an ASHA journal.

Appendix

SR #	Report Date	NTIS #	ERIC #
SR-21/22	January-June 1970	AD 719382	ED 044-679
SR-23	July-September 1970	AD 723586	ED 052-654
SR-24	October-December 1970	AD 727616	ED 052-653
SR-25/26	January-June 1971	AD 730013	ED 056-560
SR-27	July-September 1971	AD 749339	ED 071-533
SR-28	October-December 1971	AD 742140	ED 061-837
SR-29/30	January-June 1972	AD 750001	ED 071-484
SR-31/32	July-December 1972	AD 757954	ED 077-285
SR-33	January-March 1973	AD 762373	ED 081-263
SR-34	April-June 1973	AD 766178	ED 081-295
SR-35/36	July-December 1973	AD 774799	ED 094-444
SR-37/38	January-June 1974	AD 783548	ED 094-445
SR-39/40	July-December 1974	AD A007342	ED 102-633
SR-41	January-March 1975	AD A013325	ED 109-722
SR-42/43	April-September 1975	AD A018369	ED 117-770
SR-44	October-December 1975	AD A023059	ED 119-273
SR-45/46	January-June 1976	AD A026196	ED 123-678
SR-47	July-September 1976	AD A031789	ED 128-870
SR-48	October-December 1976	AD A036735	ED 135-028
SR-49	January-March 1977	AD A041460	ED 141-864
SR-50	April-June 1977	AD A044820	ED 144-138
SR-51/52	July-December 1977	AD A049215	ED 147-892
SR-53	January-March 1978	AD A055853	ED 155-760
SR-54	April-June 1978	AD A067070	ED 161-096
SR-55/56	July-December 1978	AD A065575	ED 166-757
SR-57	January-March 1979	AD A083179	ED 170-823
SR-58	April-June 1979	AD A077663	ED 178-967
SR-59/60	July-December 1979	AD A082034	ED 181-525
SR-61	January-March 1980	AD A085320	ED 185-636
SR-62	April-June 1980	AD A095062	ED 196-099
SR-63/64	July-December 1980	AD A095860	ED 197-416
SR-65	January-March 1981	AD A099958	ED 201-022
SR-66	April-June 1981	AD A105090	ED 206-038
SR-67/68	July-December 1981	AD A111385	ED 212-010
SR-69	January-March 1982	AD A120819	ED 214-226
SR-70	April-June 1982	AD A119426	ED 219-834
SR-71/72	July-December 1982	AD A124596	ED 225-212
SR-73	January-March 1983	AD A129713	ED 229-816
SR-74/75	April-September 1983	AD A136416	ED 236-753
SR-76	October-December 1983	AD A140176	ED 241-973
SR-77/78	January-June 1984	AD A145585	ED 247-626
SR-79/80	July-December 1984	AD A151035	ED 252-907
SR-81	January-March 1985	AD A156294	ED 257-159
SR-82/83	April-September 1985	AD A165084	ED 266-508
SR-84	October-December 1985	AD A168819	ED 270-831
SR-85	January-March 1986	AD A173677	ED 274-022
SR-86/87	April-September 1986	AD A176816	ED 278-066
SR-88	October-December 1986	PB 88-244256	ED 282-278

SR-89/90	January-June 1987	PB 88-244314	ED 285-228
SR-91	July-September 1987	AD A192081	**
SR-92	October-December 1987	PB 88-246798	**
SR-93/94	January-June 1988	PB 89-108765	**
SR-95/96	July-December 1988	PB 89-155329	**
SR-97/98	January-June 1989	PB 90-121161	ED 32-1317
SR-99/100	July-December 1989	PB 90-226143	ED 32-1318
SR-101/102	January-June 1990	PB 91-138479	ED 325-897
SR-103/104	July-December 1990	PB 91-172924	ED 331-100
SR-105/106	January-June 1991	PB 92-105204	ED 340-053
SR-107/108	July-December 1991	PB 92-160522	ED 344-259
SR-109/110	January-June 1992	PB 93-142099	ED 352-594
SR-111/112	July-December 1992	PB 93-216018	ED 359-575
SR-113	January-March 1993	PB 94-147220	ED 366-020
SR-114	April-June 1993	PB 94-196136	
SR-115/116	July-December 1993		

AD numbers may be ordered from:

U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Road
Springfield, VA 22151

ED numbers may be ordered from:

ERIC Document Reproduction Service
Computer Microfilm Corporation (CMC)
3900 Wheeler Avenue
Alexandria, VA 22304-5110

In addition, Haskins Laboratories Status Report on Speech Research is abstracted in *Language and Language Behavior Abstracts*, P.O. Box 22206, San Diego, CA 92122

**Accession number not yet assigned

Haskins Laboratories Status Report on Speech Research

SR-115/116
JULY-DECEMBER 1993

Contents

- Dynamics and Coordinate Systems in Skilled Sensorimotor Activity
  Elliot L. Saltzman 1
- Speech Motor Coordination and Control: Evidence From Lip, Jaw, and Laryngeal Movements
  Vincent L. Gracco and Anders Löfqvist 17
- An Unsupervised Method for Learning to Track Tongue Position from an Acoustic Signal
  John Hogden, Philip Rubin, and Elliot Saltzman 33
- Prosodic Patterns in the Coordination of Vowel and Consonant Gestures
  Caroline L. Smith 45
- Divergent Developmental Patterns for Infants' Perception of Two Non-Native Consonant Contrasts
  Catherine T. Best, Gerald W. McRoberts, Rosemarie LaFleur, and Jean Silver-Isenstadt 57
- Beyond Orthography and Phonology: Differences between Inflections and Derivations
  Laurie Beth Feldman 69
- Visual and Phonological Determinants of Misreadings in a Transparent Orthography
  G. Cossu, D. P. Shankweiler, I. Y. Liberman, and M. Gugliotta 99
- Phonological Computation and Missing Vowels: Mapping Lexical Involvement in Reading
  Ram Frost 113
- The Tritone Paradox and the Pitch Range of the Speaking Voice: A Dubious Connection
  Bruno H. Repp 127
- A Review of Treiman, R. (1993). *Beginning to Spell*
  Donald Shankweiler 145
- A Review of McNeill, D. (1992). *Hand and Mind: What Gestures Reveal About Thought*
  Michael Studdert-Kennedy 149
- A Review of Lieberman, P. (1991). *Uniquely Human*
  Michael Studdert-Kennedy 155

Appendix 157